

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
6 June 2002 (06.06.2002)

PCT

(10) International Publication Number  
WO 02/44386 A2

(51) International Patent Classification?: C12N 15/67

(US). COLLINGWOOD, Trevor [NZ/US]; Apartment 3924, 3400 Richmond Parkway, San Pablo, CA 94806 (US).

(21) International Application Number: PCT/US01/45098

(74) Agents: PASTERNAK, Dahna, S. et al.; Robins & Pasternak LLP, Suite 180, 545 Middlefield Road, Menlo Park, CA 94025 (US).

(22) International Filing Date:  
30 November 2001 (30.11.2001)

(81) Designated States (national): AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CO, CU, CZ, DE, DK, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW.

(25) Filing Language: English

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: English

Published:

— without international search report and to be republished upon receipt of that report

(30) Priority Data:  
60/250,804 1 December 2000 (01.12.2000) US

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/250,804 (CIP)  
Filed on 1 December 2000 (01.12.2000)

(71) Applicant (for all designated States except US): SANG-AMO BIOSCIENCES, INC. [US/US]; Point Richmond Tech Center, 501 Canal Boulevard, Suite A100, Richmond, CA 94804 (US).

(72) Inventor: WOLFFE, Alan, P. (deceased).

(72) Inventors; and

(75) Inventors/Applicants (for US only): TSE, Christin [US/US]; 6705 Alta Vista Drive, El Cerrito, CA 94530



WO 02/44386 A2

(54) Title: TARGETED REGULATION OF GENE EXPRESSION

(57) Abstract: Methods and compositions are provided for targeted regulation of various genes, including activation and repression of genes encoding nuclear receptors. The ability to regulate gene expression and function will have applications in treatment of disease, for example, cancer, diabetes and cardiovascular disease.

**TARGETED REGULATION OF GENE EXPRESSION****TECHNICAL FIELD**

5        This disclosure is in the field of molecular biology and medicine. More specifically, it relates to nuclear hormone receptors, regulation of their expression, and their regulation of downstream genes.

**BACKGROUND**

10       The nuclear hormone receptor superfamily plays a vital role in many physiological functions including development, cell proliferation and differentiation, and metabolism. The classical nuclear receptors (e.g., glucocorticoid and estrogen receptors) are ligand-dependent transcription factors. Many members of the nuclear receptor superfamily have no known cognate ligand and are referred to as 'orphan' receptors. While numerous nuclear receptors have unknown functions, there are several others that have been implicated in disease states such as cancer, diabetes, and hormone resistance syndromes. Thus, there is a strong probability that the majority of nuclear receptors play vital roles in cellular homeostasis.

15       The nuclear hormone receptor (NR) superfamily consists of ~65 functionally diverse receptors that operate in either ligand-dependent or -independent fashion. This superfamily includes the classical steroid receptors (e.g., androgen, estrogen, and glucocorticoid) and non-steroid receptors (e.g., thyroid, retinoid, and 'orphan'). Nuclear receptors are essential in a plethora of biological processes including development, homeostasis, cell proliferation and differentiation, and lipid metabolism.

20       In general, nuclear receptors contain an N-terminal domain (A/B), a central zinc-finger DNA-binding domain (C), and a ligand-binding domain (D/E/F). Within the A/B region of a subset of NRs is a constitutively active activation function (AF-1) whilst the D/E/F region contains a ligand-dependent activation function (AF-2) (1). Furthermore, nuclear receptors function as homodimers, heterodimers, or monomers

25       (2).

**Nuclear Hormone Receptors in Transcriptional Repression**

Some nuclear receptors repress transcription in the absence of ligand by the recruitment of co-repressors. Two co-repressors known to directly interact with NR

and mediate repression are the Silencing Mediator for Retinoid and Thyroid receptors (SMRT) and Nuclear receptor Co-Repressor (NCoR) (26, 27). This repression is thought to be through the recruitment of Sin3-HDAC complexes (class I HDACs 1-3), which deacetylate the histone N-termini, leading to the formation of a condensed chromatin structure (28, 29). However, a recent study has shown that both co-repressors can also function in a Sin3-independent pathway through the recruitment of the class II HDACs 4 and 5 (30). Several experiments suggest that the site on the ligand-binding domain (LBD) of NR that recruits repressors overlaps with the site that recruits activators (27). Thus, the molecular 'switch' for determining the preference for co-repressors or activators is the ligand, which causes a conformational change within the LBD, weakening the contacts with co-repressors and concomitantly increasing the affinity for coactivators. Furthermore, NR release of the co-repressors and subsequent recruitment of a complex with histone acetyltransferase (HAT) activity leads to the decondensation of the chromosomal locus via acetylation of the core histone N-termini and the ensuing activation of gene expression.

*Nuclear Hormone Receptors in Transcriptional Activation*

Several nuclear receptors function as transcriptional regulators that enhance gene expression when bound to their ligand. This activation is mediated through coactivators (31). A coactivator interacts directly with the AF-1 or AF-2 region of an NR, recruits components of the basal transcription machinery, and enhances transcriptional activity in the presence of the NR (31). For example, Steroid Receptor Coactivator 1 (SRC-1) has been shown to mediate the transcriptional activation for the estrogen, glucocorticoid, progesterone, thyroid hormone, and retinoid receptors (32-36). Upon ligand binding, these NRs directly interact with SRC-1, which recruits other transcription factors including the CREB-Binding Protein (CBP)/p300, P/CAF and promotes gene activation via multiple mechanisms. Another novel mechanism of transcriptional activation by NRs has recently been identified in which a NR bound to its response element is constitutively active in the absence of its ligand (37, 38). Constitutively Active Receptor (CAR) also known as MB67 is a nuclear orphan receptor that functions as a heterodimer with RXR (38, 39). In contrast to the mechanism of classical NRs, CAR appears to elicit its transcriptional activation through the recruitment of SRC-1 in the absence of ligand. Recently, androstane

metabolites have been found to serve as a ligand for CAR- $\beta$  (23). Formation of an androstanol or androstenol-CAR- $\beta$  complex causes dissociation of the coactivators and effectively represses gene expression. In sum, transcriptional activation by NRs can occur in a ligand-dependent or independent manner.

5        Given the diverse roles of NRs, a gene tool that could regulate their expression would provide a powerful means to systematically dissect their function in a particular context. Currently, there exist technologies to overexpress a gene of interest. However, this usually involves placing the gene downstream of a generic promoter, e.g., CMV or SV40, which may express the gene at levels dissimilar to those *in vivo*.

10      With respect to repression, no simple strategies are available. Mouse knockouts, for example, provide only approximations to real human tissue. Moreover, knockouts may lead to a lethal phenotype, especially if multiple knockouts are desired. In contrast antisense can be used with human tissues, but often yields only modest repression effects (40, 41). Accordingly, there is a need for reliable methods for both

15      activating and repressing the expression of NRs.

#### SUMMARY

Disclosed herein are methods and compositions for regulation of the expression of genes. In one exemplary embodiment, the gene(s) targeted for regulation encodes a

20      nuclear hormone receptor(s). Such receptors include, but are not limited to estrogen receptor alpha (ER $\alpha$ ), estrogen receptor beta (ER $\beta$ ), hepatocyte nuclear factor 4 alpha (HNF4 $\alpha$ ), hepatocyte nuclear factor 4 gamma (HNF4 $\gamma$ ), peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ), retinoid X receptor alpha (RXR $\alpha$ ), constitutively active receptor alpha (CAR $\alpha$ ) and androgen receptor (AR).

25      The compositions include regulatory molecules comprising a DNA-binding domain (preferably a zinc finger domain) and a functional domain. The functional domain can be an activation domain or a repression domain. Exemplary activation domains include, but are not limited to, VP16, p65 or functional fragments thereof. Exemplary repression domains include, but are not limited to, KRAB, thyroid hormone

30      receptor (TR), vErbA and functional fragments thereof. Polynucleotide sequences encoding the regulatory molecules, optionally as part of an expression vector, are also provided, as are cells comprising the regulatory molecules and cells comprising polynucleotides and/or expression vectors encoding the regulatory molecules.

Methods for regulation of genes (e.g., NR genes) comprise identifying one or more accessible regions in cellular chromatin comprising the gene, examining the nucleotide sequence of the accessible region(s), and designing the DNA-binding domain of the regulatory molecule to target a sequence within an accessible region.

5 The regulatory molecule so designed, comprising a DNA-binding domain targeted to a sequence in an accessible region of the gene, and either an activation or a repression domain, is contacted with the cell. Alternatively, or in addition, a polynucleotide encoding the regulatory molecule (optionally contained in an expression construct) is contacted with the cell. Modulation of expression of the gene of interest is assayed by  
10 standard methods, such as measurements of RNA (TaqMan, RNA blot, RNase protection) or protein (ELISA, protein immunoblot) levels.

Modulation of expression of target genes can also result in modulation of expression of additional genes. For example, in embodiments in which the target gene is a nuclear receptor, the methods can also result in modulation of expression of genes  
15 whose expression is regulated by the NR. Thus, the disclosure also provides methods and compositions for modulation of expression of genes whose expression is regulated by a target gene such as a NR gene.

#### BRIEF DESCRIPTION OF THE FIGURES

20 Figure 1 depicts interactions between Zif268, a canonical three-finger DNA-binding domain, and a ten base-pair DNA target.

Figure 2, panel A is a schematic diagram of the promoter region of Estrogen Receptor Alpha. Figure 2, panel B shows a gel of the DNase I mapping (HS designates hypersensitive site). The two cell lines mapped are MDA-MB-231  
25 that does not express any detectable mRNA and MCF-7 that expresses appreciable levels of mRNA. Both cell lines are breast cancer-derived. Shown on the diagram are the restriction sites: Xb = Xba I, B = Bam HI, EV = Eco RV, EI = Eco RI, Xm = Xma I. The Probe designated by the hashed box on the 5' region of the promoter. Two transcription start sites have been identified. P1 is the primary  
30 promoter utilized.

Figure 3 shows mapping of DNase I hypersensitive sites (HS) of PPAR- $\gamma$ 2 promoter. Engineered ZFPs (zfp52, zfp54, and zfp55) were designed around the vicinity of HS1 near the transcriptional start site (promoter B). A second

transcription start site is located several kilobases upstream of the one shown in this diagram. The probe was designed to recognize the 5' end of the promoter.

Figure 4 shows real-time PCR of total RNA isolated from 3T3-L1 fibroblast for gene expression of PPAR- $\gamma$ 1 (Promoter A) and PPAR- $\gamma$ 2 (Promoter B).

Figure 5, panels A and B are schematic diagrams of the genomic walking protocol (Clontech) and 5' RACE (Ambion), adapted from GenomeWalker<sup>TM</sup> and RLM-RACE User Manuals. AP1, 2 = adapter primers; GSP1,2 = gene specific primers; CIP = calf intestinal phosphatase; TAP = tobacco acid pyrophosphatase.

10 Figure 6 is a graph depicting repression of ER- $\alpha$  in MCF-7 cells.

Figure 7 is a schematic diagram of ZFP target sites for ER- $\alpha$  activation. DNAse hypersites identified in ER(+) breast carcinoma cell lines represent important regulatory sequences at -3810, -2100 and -320.

Figure 8 is a graph depicting activation of ER- $\alpha$  with functional domains.

15 The gray, black and white bars show mRNA expression at .3  $\mu$ g, .6  $\mu$ g and .9  $\mu$ g concentrations, respectively.

#### DETAILED DESCRIPTION

The practice of the disclosed methods and use of the disclosed compositions 20 employ, unless otherwise indicated, conventional techniques in molecular biology, biochemistry, genetics, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. *See, for example, Sambrook et al. MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; and the series METHODS IN ENZYMOLOGY, Academic Press, San Diego.*

The disclosures of all patents, patent applications and publications mentioned herein are hereby incorporated by reference in their entireties.

30

#### **Definitions**

The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer in either

single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties. In general, an analogue 5 of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analogue of A will base-pair with T. Thus, the term polynucleotide sequence is the alphabetical representation of a polynucleotide molecule. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

10 Chromatin is the nucleoprotein structure comprising the cellular genome. “Cellular chromatin” comprises nucleic acid, primarily DNA, and protein, including histones and non-histone chromosomal proteins. The majority of eukaryotic cellular chromatin exists in the form of nucleosomes, wherein a nucleosome core comprises approximately 150 base pairs of DNA associated with 15 an octamer comprising two each of histones H2A, H2B, H3 and H4; and linker DNA (of variable length depending on the organism) extends between nucleosome cores. A molecule of histone H1 is generally associated with the linker DNA. For the purposes of the present disclosure, the term “chromatin” is meant to encompass all types of cellular nucleoprotein, both prokaryotic and 20 eukaryotic. Cellular chromatin includes both chromosomal and episomal chromatin.

A “chromosome” is a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype, which is the collection of all the chromosomes that comprise the genome of the 25 cell. The genome of a cell can comprise one or more chromosomes.

An “episome” is a replicating nucleic acid, nucleoprotein complex or other structure comprising a nucleic acid that is not part of the chromosomal karyotype of a cell. Examples of episomes include plasmids and certain viral genomes.

Typical “control elements” include, but are not limited to, transcription 30 promoters, transcription enhancer elements, cis-acting transcription regulating elements (transcription regulators, *e.g.*, a cis-acting element that affects the transcription of a gene, for example, a region of a promoter with which a transcription factor interacts to modulate expression of a gene), transcription termination signals, as well as polyadenylation sequences (located 3' to the translation stop codon), sequences

for optimization of initiation of translation (located 5' to the coding sequence), translation enhancing sequences, and translation termination sequences. Control elements are preferably derived from the polynucleotides described herein (e.g., NR sequences) and include functional fragments thereof, for example, polynucleotides 5 between about 5 and about 50 nucleotides in length (or any integer therebetween); preferably between about 5 and about 25 nucleotides (or any integer therebetween), even more preferably between about 5 and about 10 nucleotides (or any integer therebetween), and most preferably 9-10 nucleotides. Transcription promoters can include inducible promoters (where expression of a polynucleotide sequence operably 10 linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), repressible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), and constitutive promoters.

Techniques for determining nucleic acid and amino acid "sequence 15 identity" also are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic sequences can also be determined and compared in this fashion. In general, "identity" refers to an exact 20 nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences 25 divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and 30 Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, WI) in the "BestFit" utility application. The default parameters

for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). A preferred method of establishing percent identity in the context of the present disclosure is to use the MPSRCH package of programs copyrighted 5 by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated the 10 "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code = standard; filter = none; strand = 15 both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at the following internet address: <http://www.ncbi.nlm.gov/cgi-bin/BLAST>. When claiming sequences relative to 20 sequences described herein, the range of desired degrees of sequence identity is approximately 80% to 100% and any integer value therebetween. Typically the percent identities between the disclosed sequences and the claimed sequences are at least 70-75%, preferably 80-82%, more preferably 85-90%, even more preferably 92%, still more preferably 95%, and most preferably 98% sequence 25 identity to the reference sequence (i.e., the sequences disclosed herein).

Alternatively, the degree of sequence similarity between polynucleotides can be determined by hybridization of polynucleotides under conditions that allow formation of stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested 30 fragments. Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 70%-75%, preferably 80%-82%, more preferably 85%-90%, even more preferably 92%, still more preferably 95%, and most preferably 98% sequence identity to the reference sequence over a defined length of the molecules, as determined using the methods

above. As used herein, substantially homologous also refers to sequences showing complete identity to the specified DNA or polypeptide sequence. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for 5 that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., *supra*; DNA Cloning: A Practical Approach, editor, D.M. Glover (1985) Oxford; Washington, DC; IRL Press; Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins (1985) Oxford; Washington, DC; IRL Press.

10 "Selective hybridization" of two nucleic acid fragments can be determined as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit the hybridization of a completely identical sequence to a target molecule.

15 Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for 20 example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

25 When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A 30 nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" hybridization conditions typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent

hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe.

Hybridization conditions useful for probe/target hybridization where the probe  
5 and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

Conditions for hybridization are well-known to those of skill in the art.  
10 Hybridization stringency refers to the degree to which hybridization conditions disfavor the formation of hybrids containing mismatched nucleotides, with higher stringency correlated with a lower tolerance for mismatched hybrids. Factors that affect the stringency of hybridization are well-known to those of skill in the art and include, but are not limited to, temperature, pH, ionic strength, and  
15 concentration of organic solvents such as, for example, formamide and dimethylsulfoxide. As is known to those of skill in the art, hybridization stringency is increased by higher temperatures, lower ionic strength and lower solvent concentrations.

With respect to stringency conditions for hybridization, it is well known in the  
20 art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., dextran sulfate, and polyethylene  
25 glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

30 The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of corresponding naturally-occurring amino acids.

A “binding protein” is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can 5 bind to itself (to form homodimers, homotrimers, *etc.*) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

10 A “zinc finger DNA binding protein” is a protein or segment within a larger protein that binds DNA in a sequence-specific manner as a result of stabilization of protein structure through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

15 A “designed” zinc finger protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP designs and binding data. A “selected” zinc finger protein is a protein not found in nature whose production results primarily from an empirical process such as phage display. See *e.g.*, US 5,789,538; US 6,007,988; US 6,013,453; US 6,140,081; US 6,140,466; 20 WO 95/19431; WO 96/06166 and WO 98/54311.

The term “naturally-occurring” is used to describe an object that can be found in nature, as distinct from being artificially produced by humans.

25 Nucleic acid or amino acid sequences are “operably linked” (or “operatively linked”) when placed into a functional relationship with one another. For instance, a promoter or enhancer is operably linked to a coding sequence if it regulates, or contributes to the modulation of, the transcription of the coding sequence. Operably linked DNA sequences are typically joined in *cis* and can be contiguous, and operably linked amino acid sequences are typically contiguous and in the same reading frame. However, since enhancers generally function when separated from the promoter by up 30 to several kilobases or more and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous. Similarly, certain amino acid sequences that are non-contiguous in a primary polypeptide sequence may nonetheless be operably linked due to, for example folding of a polypeptide chain.

With respect to fusion polypeptides, the term "operatively linked" can refer to the fact that each of the components performs the same function in linkage to the other component as it would if it were not so linked. For example, with respect to a fusion polypeptide in which a ZFP DNA-binding domain is fused to a transcriptional

5 activation domain (or functional fragment thereof), the ZFP DNA-binding domain and the transcriptional activation domain (or functional fragment thereof) are in operative linkage if, in the fusion polypeptide, the ZFP DNA-binding domain portion is able to bind its target site and/or its binding site, while the transcriptional activation domain (or functional fragment thereof) is able to activate transcription.

10 A "functional fragment" of a protein, polypeptide or nucleic acid is a protein, polypeptide or nucleic acid whose sequence is not identical to the full-length protein, polypeptide or nucleic acid, yet retains the same function as the full-length protein, polypeptide or nucleic acid. A functional fragment can possess more, fewer, or the same number of residues as the corresponding native

15 molecule, and/or can contain one or more amino acid or nucleotide substitutions. Methods for determining the function of a nucleic acid (e.g., coding function, ability to hybridize to another nucleic acid, binding to a regulatory molecule) are well-known in the art. Similarly, methods for determining protein function are well-known. For example, the DNA-binding function of a polypeptide can be

20 determined, for example, by filter-binding, electrophoretic mobility-shift, or immunoprecipitation assays. See Ausubel *et al.*, *supra*. The ability of a protein to interact with another protein can be determined, for example, by co-immunoprecipitation, two-hybrid assays or complementation, both genetic and biochemical. See, for example, Fields *et al.* (1989) *Nature* 340:245-246; U.S.

25 Patent No. 5,585,245 and PCT WO 98/44350.

"Specific binding" between, for example, a ZFP and a specific target site means a binding affinity of at least  $1 \times 10^6 \text{ M}^{-1}$ .

A "fusion molecule" is a molecule in which two or more subunit molecules are linked, preferably covalently. The subunit molecules can be the same chemical type of molecule, or can be different chemical types of molecules. Examples of the first type of fusion molecule include, but are not limited to, fusion polypeptides (for example, a fusion between a ZFP DNA-binding domain and a methyl binding domain) and fusion nucleic acids (for example, a nucleic acid encoding a fusion polypeptide). Examples of the second type of fusion molecule include, but are not limited to, a fusion between

a triplex-forming nucleic acid and a polypeptide, and a fusion between a minor groove binder and a nucleic acid.

An “exogenous molecule” is a molecule that is not normally present in a cell, but can be introduced into a cell by one or more genetic, biochemical or other methods. Normal presence in the cell is determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell. An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule or a malfunctioning version of a normally-functioning endogenous molecule.

An exogenous molecule can be, among other things, a small molecule, such as is generated by a combinatorial chemistry process, or a macromolecule such as a protein, nucleic acid, carbohydrate, lipid, glycoprotein, lipoprotein, polysaccharide, any modified derivative of the above molecules, or any complex comprising one or more of the above molecules. Nucleic acids include DNA and RNA, can be single- or double-stranded; can be linear, branched or circular; and can be of any length. Nucleic acids include those capable of forming duplexes, as well as triplex-forming nucleic acids. See, for example, U.S. Patent Nos. 5,176,996 and 5,422,251. Proteins include, but are not limited to, DNA-binding proteins, transcription factors, chromatin remodeling factors, methylated DNA binding proteins, polymerases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, integrases, recombinases, ligases, topoisomerases, gyrases and helicases.

An exogenous molecule can be the same type of molecule as an endogenous molecule, *e.g.*, protein or nucleic acid (*i.e.*, an exogenous gene), providing it has a sequence that is different from an endogenous molecule. For example, an exogenous nucleic acid can comprise an infecting viral genome, a plasmid or episome introduced into a cell, or a chromosome that is not normally present in the cell. Methods for the introduction of exogenous molecules into cells are known to those of skill in the art and include, but are not limited to, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment, calcium

phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer.

By contrast, an “endogenous molecule” is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally-occurring episomal nucleic acid. Additional endogenous molecules can include endogenous genes and endogenous proteins, for example, transcription factors and components of chromatin remodeling complexes.

10 A “gene,” for the purposes of the present disclosure, includes a DNA region encoding a gene product (see below), as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

20 “Gene expression” refers to the conversion of the information, contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (e.g., mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of a mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and 25 glycosylation.

“Gene activation” and “augmentation of gene expression” refer to any process which results in an increase in production of a gene product. A gene product can be either RNA (including, but not limited to, mRNA, rRNA, tRNA, and structural RNA) or protein. Accordingly, gene activation includes those processes which increase 30 transcription of a gene and/or translation of a mRNA. Examples of gene activation processes which increase transcription include, but are not limited to, those which facilitate formation of a transcription initiation complex, those which increase transcription initiation rate, those which increase transcription elongation rate, those which increase processivity of transcription and those which relieve transcriptional

repression (by, for example, blocking the binding of a transcriptional repressor). Gene activation can constitute, for example, inhibition of repression as well as stimulation of expression above an existing level. Examples of gene activation processes which increase translation include those which increase translational initiation, those which 5 increase translational elongation and those which increase mRNA stability. In general, gene activation comprises any detectable increase in the production of a gene product, preferably an increase in production of a gene product by about 2-fold, more preferably from about 2- to about 5-fold or any integral value therebetween, more preferably between about 5- and about 10-fold or any integral value therebetween, more 10 preferably between about 10- and about 20-fold or any integral value therebetween, still more preferably between about 20- and about 50-fold or any integral value therebetween, more preferably between about 50- and about 100-fold or any integral value therebetween, more preferably 100-fold or more.

“Gene repression” and “inhibition of gene expression” refer to any process 15 which results in a decrease in production of a gene product. A gene product can be either RNA (including, but not limited to, mRNA, rRNA, tRNA, and structural RNA) or protein. Accordingly, gene repression includes those processes which decrease transcription of a gene and/or translation of a mRNA. Examples of gene repression processes which decrease transcription include, but are not limited to, those which 20 inhibit formation of a transcription initiation complex, those which decrease transcription initiation rate, those which decrease transcription elongation rate, those which decrease processivity of transcription and those which antagonize transcriptional activation (by, for example, blocking the binding of a transcriptional activator). Gene repression can constitute, for example, prevention of activation as well as inhibition of 25 expression below an existing level. Examples of gene repression processes which decrease translation include those which decrease translational initiation, those which decrease translational elongation and those which decrease mRNA stability.

Transcriptional repression includes both reversible and irreversible inactivation of gene transcription. In general, gene repression comprises any detectable decrease in the 30 production of a gene product, preferably a decrease in production of a gene product by about 2-fold, more preferably from about 2- to about 5-fold or any integral value therebetween, more preferably between about 5- and about 10-fold or any integral value therebetween, more preferably between about 10- and about 20-fold or any integral value therebetween, still more preferably between about 20- and about 50-fold

or any integral value therebetween, more preferably between about 50- and about 100-fold or any integral value therebetween, more preferably 100-fold or more. Most preferably, gene repression results in complete inhibition of gene expression, such that no gene product is detectable.

5       “Modulation” of gene expression includes both gene activation and gene repression. Modulation can be assayed by determining any parameter that is indirectly or directly affected by the expression of the target gene. Such parameters include, e.g., changes in RNA or protein levels; changes in protein activity; changes in product levels; changes in downstream gene expression; changes in transcription or activity of 10 reporter genes such as, for example, luciferase, CAT, beta-galactosidase, or GFP (see, e.g., Mistili & Spector, (1997) *Nature Biotechnology* 15:961-964); changes in signal transduction; changes in phosphorylation and dephosphorylation; changes in receptor-ligand interactions; changes in concentrations of second messengers such as, for example, cGMP, cAMP, IP<sub>3</sub>, and Ca<sup>2+</sup>; changes in cell growth, changes in 15 neovascularization, and/or changes in any functional effect of gene expression. Measurements can be made *in vitro*, *in vivo*, and/or *ex vivo*. Such functional effects can be measured by conventional methods, e.g., measurement of RNA or protein levels, measurement of RNA stability, and/or identification of downstream or reporter gene expression. Readout can be by way of, for example, chemiluminescence, 20 fluorescence, colorimetric reactions, antibody binding, inducible markers, ligand binding assays; changes in intracellular second messengers such as cGMP and inositol triphosphate (IP<sub>3</sub>); changes in intracellular calcium levels; cytokine release, and the like.

“Eucaryotic cells” include, but are not limited to, fungal cells (such as yeast), 25 plant cells, animal cells, mammalian cells and human cells.

A “regulatory domain” or “functional domain” refers to a protein or a polypeptide sequence that has transcriptional modulation activity. In one embodiment, a regulatory domain is covalently or non-covalently linked to a ZFP to modulate transcription of a gene of interest. Alternatively, a ZFP can act alone, without a 30 regulatory domain, to modulate transcription. Furthermore, transcription of a gene of interest can be modulated by a ZFP linked to multiple regulatory domains. In addition, a regulatory domain can be linked to any DNA-binding domain having the appropriate specificity to modulate the expression of a gene of interest.

A “target site” or “target sequence” is a sequence that is bound by a binding protein or binding domain such as, for example, a ZFP. Target sequences can be nucleotide sequences (either DNA or RNA) or amino acid sequences. By way of example, a DNA target sequence for a three-finger ZFP is generally either 9 or 10 5 nucleotides in length, depending upon the presence and/or nature of cross-strand interactions between the ZFP and the target sequence.

The term “recombinant,” when used with reference to a cell, indicates that the cell replicates an exogenous nucleic acid, or expresses a peptide or protein 10 encoded by an exogenous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) form of the cell. Recombinant 15 cells can also contain genes found in the native form of the cell wherein the genes are modified and re-introduced into the cell. The term also encompasses cells that contain a nucleic acid endogenous to the cell that has been modified without removing the nucleic acid from the cell; such modifications include those obtained by gene replacement, site-specific mutation, and related techniques.

A “recombinant expression cassette,” “expression cassette” or “expression construct” is a nucleic acid construct, generated recombinantly or synthetically, 20 that has control elements that are capable of effecting expression of a structural gene that is operatively linked to the control elements in hosts compatible with such sequences. Expression cassettes include at least promoters and optionally, transcription termination signals. Typically, the recombinant expression cassette includes at least a nucleic acid to be transcribed (e.g., a nucleic acid encoding a 25 desired polypeptide) and a promoter. Additional factors necessary or helpful in effecting expression can also be used as described herein. For example, an expression cassette can also include nucleotide sequences that encode a signal sequence that directs secretion of an expressed protein from the host cell, nuclear localization signals and/or epitope tags. Transcription termination signals, enhancers, and other nucleic acid sequences that influence gene expression, can also be included in an expression cassette.

30

### Overview

The compositions and methods disclosed herein allow for targeted regulation of genes, for example targeted regulation of genes encoding various nuclear hormone receptors (NRs). Regulation includes modulation of gene

expression, which includes activation and repression of gene expression. The effects of increased or decreased expression of a particular gene can be assessed, for example, by changes in patterns of cellular transcription which accompany modulation of gene expression. Regulation of gene expression (such as nuclear 5 receptor genes) will be useful in treatment of various diseases, including cancer, diabetes and cardiovascular disease.

### Target Nuclear Receptors

In certain embodiments, the gene regulated by the methods described 10 herein encodes a nuclear receptor. Table 1 lists the initial nuclear receptor targets. To elucidate the functions of all the nuclear receptors, it is first necessary to test nuclear receptors of known function. In this regard, a fair amount of biology is known for AR, ER- $\alpha$ , PPAR- $\gamma$ , and RXR- $\alpha$ , which will allow for direct comparison of results. Secondly, gene knockouts are available for each of these 15 targets, except the androgen receptor (3-10). These provide a point of reference for comparison of results obtained using the methods and compositions disclosed herein for down-regulation of NR expression, particularly in studies involving transgenic mice. Seven out of eight of the target genes are known to play an essential role in various disease states. For example, HNF4- $\alpha,\gamma$  are essential for 20 glucose, cholesterol, and fatty acid metabolism (7). Defects in the pathway lead to type I diabetes (11, 12). PPAR- $\gamma$  is also involved in glucose metabolism and appears to be misregulated in type II diabetes (13). In fact, PPAR- $\gamma$  is a primary pharmacologic target in the treatment of type II diabetes (14). Both ER- $\alpha,\beta$  are 25 involved in several cancers. ER- $\alpha$  is known to be directly associated with breast carcinomas (15-17). Approximately 50% of all breast cancers express unusually high ER- $\alpha$  levels. Hence, ER- $\alpha$  has become a primary target for anti-cancer agents (17). RXR- $\alpha$  can function as a homodimer and as a heterodimer with the RAR (retinoic acid receptor), PPAR- $\gamma$ , CAR- $\beta$ , and many other receptors (18-22). As such, RXR- $\alpha$  is a common regulatory component of multiple physiologic 30 pathways and is involved in several disease states. CAR- $\beta$  is thought to be involved in toxin responses (4, 23); and understanding of the functions of this nuclear receptor will be expanded using the compositions and methods disclosed herein. Finally, AR is involved in normal male sexual differentiation; defects in

AR expression and/or regulation lead to androgen insensitivity syndrome (24). Furthermore, somatic mutations in AR have been associated with prostate cancer (25). In sum, the compositions and methods disclosed herein can be used to regulate these NR targets. Furthermore, elucidation of the function of additional 5 members of the nuclear receptor superfamily can also be achieved suing the methods and compositions disclosed herein.

Table 1

Trivial Name	Gene	Disease	Genomic Sequence
HNF4 $\alpha$	NR2A1	DM	p/n
HNF4 $\gamma$	N2AB2	DM	p/n
PPAR $\gamma$	NR1C3	C, DM	p/p
RXR $\alpha$	NR2B1	S	n/n
CAR $\alpha$	NR1I3	U	n/n
ER $\alpha$	NR3A1	C	w/n
ER $\beta$	NR3A2	C	w/n
AR	NR3C4	AIS, FX, O	w/p

Table 1. Nuclear hormone receptor targets: HNF = Hepatocyte Nuclear Factor; PPAR = Peroxisome Proliferator-Activated Receptor; RXR = Retinoid X Receptor; CAR = Constitutively Active Receptor; ER = Estrogen Receptor; AR = Androgen Receptor; NR = nuclear hormone receptor. Disease states: AIS = androgen insensitivity syndrome; C = cancer; D = diabetes mellitus; FX = Fragile X syndrome; S = several; U = unknown; O = others e.g., breast and prostate cancer and muscular atrophy. Genomic sequence availability in human / mouse, w = at least 2.5kb of sequence upstream of sequence is available; p = partial promoter sequence is available; n = no promoter sequence is available.

### Zinc Finger Protein (ZFP) Technology

Zinc fingers are the natural constituents of many cellular transcription factors. To date, over 30,000 zinc finger sequences have been identified in 20 thousands of known or putative transcription factors. Zinc fingers are present in arrays involved in binding specific DNA sequences, amino acid sequences, RNA helices and possibly RNA-DNA heteroduplexes during transcription initiation (42). Approximately 3% of all human genes are believed to comprise zinc finger domains (43, 44). Thus, zinc fingers are a predominant means of regulating gene 25 expression inside a cell. In general, ZFP transcription factors have two distinct

domains: (1) a DNA binding domain (DBD) that directs the ZFP to the proper chromosomal location by recognizing a specific DNA sequence and (2) a functional domain that regulates gene expression of a specific locus. This two-component structure is the foundation for the design of ZFPs to regulate NRs as 5 disclosed herein. Designed ZFPs are well-suited for targeting specific genes due to their established DNA-specificity. The validity of using designed ZFPs for regulation of endogenous gene loci has been demonstrated (45-49).

Zinc finger proteins can easily be identified according to a conserved zinc-chelating sequence, -Cys-(X)<sub>2-4</sub>-Cys-(X)<sub>3</sub>-Phe-(X)<sub>5</sub>-Leu-(X)<sub>2</sub>-His-(X)<sub>3-5</sub>-His (51). 10 A single finger domain is 30 amino acids in length and consists of two  $\beta$ -strands and an  $\alpha$ -helix containing two invariant histidine residues (52). The  $\beta$ -strands position the  $\alpha$ -helix to recognize the major groove of DNA. Zinc fingers interact with DNA as independent modules that bind preferentially in the DNA major groove. When linked together, zinc fingers can be used to target a protein to a 15 specific chromosomal locus. Zinc fingers bind their target sequence in a modular fashion, with individual fingers in a multi-fingered domain binding in the DNA major groove over three base pair intervals, as first characterized by x-ray crystallography (53-56) (Figure 1). The base-specific DNA contacts are made by the side-chains on each finger recognition helix, interacting directly with 20 functional groups of the bases within the DNA major groove.

Mutagenesis experiments have shown that it is possible to predictably alter the DNA-binding preferences of zinc fingers by making changes in the amino acid sequences of the recognition helices (57-66). Only a few side chain substitutions are required to change the DNA-binding specificity of a ZFP and if the changes 25 are limited to the same four locations on each recognition helix, the DNA-binding domain can be rationally altered to specifically bind a large combination of sequences. See, for example, co-owned WO 00/42219; WO 00/41566; and U.S. Serial Nos. 09/444,241 filed November 19, 1999; 09/535,088 filed March 23, 2000; as well as U.S. Patents 5,789,538; 6,007,408; 6,013,453; 6,140,081; 30 and 6,140,466; and PCT publications WO 95/19431, WO 98/54311, WO 00/23464 and WO 00/27878. See also Wolfe *et al.* (2000) *Ann. Rev. Biophys. Biomol. Struct.* 3:183-212 and Joung *et al.* (2000) *Proc. Natl. Acad. Sci. USA* 97:7382-7387. In one embodiment, a target site for a zinc finger DNA-binding domain is identified according to site selection rules disclosed in co-

owned WO 00/42219. In a preferred embodiment, a ZFP is selected as described in co-owned U.S. Serial No. Unassigned, filed November 20, 2000, titled "Iterative Optimization in the Design of Binding Proteins."

5   **Chromosomal Regulation of Nuclear Receptors using ZFPs linked to Various Functional Domains**

The packaging of DNA into chromatin presents a major obstacle to gene expression. Numerous studies have demonstrated the refractive nature of chromatin on transcription factor access (68-70). Thus, in certain cases, 10 transcription factors, including designed ZFPs must first gain access to their target sequences in cellular chromatin to elicit their effects. In order to rationally design a ZFP for regulation of a gene of interest, the chromatin structure of that gene's promoter is mapped to determine the 'hypersensitive' regions. Analysis of chromatin structure allows the identification of potential regulatory sequences, 15 facilitates design of ZFPs to overcome refractory effects of chromatin structure, and defines the physiological state of the target promoter. Hence, it is essential to characterize the chromatin structure of the promoter regions of target genes. Accordingly, low- and high-resolution DNase I hypersensitive mapping techniques are employed to identify regions of the promoter that are accessible to 20 engineered ZFPs. Methods for identifying and characterizing accessible regions in cellular chromatin, using DNase hypersensitivity and other techniques, are disclosed in co-owned U.S. Patent Application Serial No. 60/228,556, entitled "Databases of Regulatory Sequences; Methods of Making and Using Same," filed August 28, 2000. See Examples 1 and 2, *infra*.

25   Thus, it is useful to characterize the chromatin structure of the promoter regions of the selected genes (e.g., NRs in Table 1) in both *Homo sapiens* and *Mus musculus*. This characterization will allow design of ZFPs that recognize accessible regions of the promoter. In particular, characterization of accessible regions in cells expression high levels of a gene product, and comparison to cells 30 expressing low levels, leads to identification of accessible regions important for regulation of gene expression.

Molecules for regulating gene expression comprise a DNA-binding domain, preferably targeted to a sequence in an accessible region of the target gene, and a functional domain. In a preferred embodiment, the DNA-binding

domain comprises one or more zinc finger domains. A functional domain can be either an activation domain or a repression domain. VP16 and p65 are preferred transcriptional activation domains. VP16 is a very potent activator that has been utilized to activate a wide range of genes (71-73). Preferred repression domains

5 include the KRAB domain and vErbA. The 90 amino acid KRuppel-Associated Box (KRAB) repressor domain is prevalent in many natural transcriptional repressors (74, 75). Another useful repression domain is that associated with the v-ErbA protein. See, for example, Damm, et al. (1989) *Nature* 339:593-597; Evans (1989) *Int. J. Cancer Suppl.* 4:26-28; Pain et al. (1990) *New Biol.* 2:284-10 294; Sap et al. (1989) *Nature* 340:242-244; Zenke et al. (1988) *Cell* 52:107-119; and Zenke et al. (1990) *Cell* 61:1035-1049. Other useful repression domains include the Methyl Binding Domains 2 and 3, DNA Methyltransferase 1, and Thyroid Hormone Receptor (TR).

To regulate a receptor gene in a living cell, a regulatory molecule, as 15 described above, is contacted with the cell. Alternatively, the cell can be contacted with a nucleic acid encoding a regulatory molecule. See *infra* for further details.

#### Applications of Methods for Regulating Nuclear Receptors

20 Nuclear hormone receptors play a vital role in a plethora of physiological pathways. They have been implicated in disease states such as cancer, acute promyelocytic leukemia (76-78), diabetes mellitus (79, 80), and hormone resistance syndromes (76-78). Furthermore, there are several nuclear receptors with unknown function. Thus, the ability to regulate gene expression of the nuclear receptor 25 superfamily is highly valuable in pharmaceutical research of both nuclear receptors of known function as well as those of unknown function. Such regulation would facilitate the development of tissue and animal models of disease states, drug validation, and therapeutic product development. Nuclear receptor regulation packages can be designed according to NR class and/or disease states. For example, an estrogen 30 receptor-  $\alpha, \beta$  package would be useful for investigating potential treatments for breast cancer. This methods and compositions disclosed herein are adaptable to complement transgenic mouse models of various diseases, by effectively creating traditional 'knockout' mice without the tedious procedures of deleting the two copies of the

endogenous gene, and by providing a means to produce inducible 'knockout' mice. These advances will facilitate production of animal models of diseases. With respects to drug validation, if a drug is developed to act as an antagonist for a particular NR, the resulting phenotype of a cell, that has been treated with the drug, can be compared to a 5 cell in which a repressing ZFP has been introduced. Hormone resistance syndromes, which arise from the premature inactivation of a nuclear receptor, can be treated by reactivating the endogenous gene. Finally, regulation of nuclear receptors of unknown function will allow identification their role(s) in cellular homeostasis.

10           **DNA-Binding domains**

In preferred embodiments, the compositions and methods disclosed herein involve use of DNA binding proteins, particular zinc finger proteins. A DNA-binding domain can comprise any molecular entity capable of sequence-specific binding to chromosomal DNA. Binding can be mediated by electrostatic 15 interactions, hydrophobic interactions, or any other type of chemical interaction. Examples of moieties which can comprise part of a DNA-binding domain include, but are not limited to, minor groove binders, major groove binders, antibiotics, intercalating agents, peptides, polypeptides, oligonucleotides, and nucleic acids. An example of a DNA-binding nucleic acid is a triplex-forming oligonucleotide.

20           Minor groove binders include substances which, by virtue of their steric and/or electrostatic properties, interact preferentially with the minor groove of double-stranded nucleic acids. Certain minor groove binders exhibit a preference for particular sequence compositions. For instance, netropsin, distamycin and CC-1065 are examples of minor groove binders which bind specifically to AT-rich sequences, particularly runs of A or T. WO 96/32496.

25           Many antibiotics are known to exert their effects by binding to DNA. Binding of antibiotics to DNA is often sequence-specific or exhibits sequence preferences. Actinomycin, for instance, is a relatively GC-specific DNA binding agent.

30           In a preferred embodiment, a DNA-binding domain is a polypeptide. Certain peptide and polypeptide sequences bind to double-stranded DNA in a sequence-specific manner. For example, transcription factors participate in transcription initiation by RNA Polymerase II through sequence-specific interactions with DNA in the promoter and/or enhancer regions of genes. Defined

regions within the polypeptide sequence of various transcription factors have been shown to be responsible for sequence-specific binding to DNA. See, for example, Pabo *et al.* (1992) *Ann. Rev. Biochem.* 61:1053-1095 and references cited therein. These regions include, but are not limited to, motifs known as leucine zippers, 5 helix-loop-helix (HLH) domains, helix-turn-helix domains, zinc fingers,  $\beta$ -sheet motifs, steroid receptor motifs, bZIP domains homeodomains, AT-hooks and others. The amino acid sequences of these motifs are known and, in some cases, amino acids that are critical for sequence specificity have been identified. Polypeptides involved in other process involving DNA, such as replication, 10 recombination and repair, will also have regions involved in specific interactions with DNA. Peptide sequences involved in specific DNA recognition, such as those found in transcription factors, can be obtained through recombinant DNA cloning and expression techniques or by chemical synthesis, and can be attached to other components of a fusion molecule by methods known in the art.

15 In a more preferred embodiment, a DNA-binding domain comprises a zinc finger DNA-binding domain. See, for example, Miller *et al.* (1985) *EMBO J.* 4:1609-1614; Rhodes *et al.* (1993) *Scientific American* Feb.:56-65; and Klug (1999) *J. Mol. Biol.* 293:215-218. The three-fingered Zif268 murine transcription factor has been particularly well studied. (Pavletich, N. P. & Pabo, C. O. (1991) *Science* 252:809-17). The X-ray co-crystal structure of Zif268 ZFP and double-stranded DNA indicates that each finger interacts independently with DNA. (Nolte *et al.* (1998) *Proc Natl Acad Sci U S A* 95:2938-43; Pavletich, N. P. & Pabo, C. O. (1993) *Science* 261:1701-7). The organization of the 3-fingered domain allows 20 recognition of three contiguous base-pair triplets by each finger. Each finger is approximately 30 amino acids long, adopting a  $\beta\beta\alpha$  fold. The two  $\beta$ -strands form a sheet, positioning the recognition  $\alpha$ -helix in the major groove for DNA binding. Specific contacts with the bases are mediated primarily by four amino acids 25 immediately preceding and within the recognition helix. Conventionally, these recognition residues are numbered -1, 2, 3, and 6 based on their positions in the  $\alpha$ -helix.

30 ZFP DNA-binding domains are designed and/or selected to recognize a particular target site as described in co-owned WO 00/42219; WO 00/41566; and U.S. Serial Nos. 09/444,241 filed November 19, 1999; 09/535,088 filed March 23,

2000; as well as U.S. Patents 5,789,538; 6,007,408; 6,013,453; 6,140,081; and 6,140,466; and PCT publications WO 95/19431, WO 98/54311, WO 00/23464 and WO 00/27878. In one embodiment, a target site for a zinc finger DNA-binding domain is identified according to site selection rules 5 disclosed in co-owned WO 00/42219. In a preferred embodiment, a ZFP is selected as described in co-owned U.S. Serial No. Unassigned, filed November 20, 2000, titled "Iterative Optimization in the Design of Binding Proteins."

In certain preferred embodiments, the binding specificity of the DNA-binding domain can be determined by identifying accessible regions in the 10 sequence in question (e.g., in cellular chromatin). Accessible regions can be determined as described in co-owned U.S. Patent Application Serial No. 60/228,556 entitled "Databases of Accessible Region Sequences; Methods of Preparation and Use Thereof," filed August 28, 2000, the disclosure of which is hereby incorporated by reference herein. See also Example 2. A DNA-binding 15 domain is then designed and/or selected as described herein to bind to a target site within the accessible region.

#### **Fusion Molecules**

The identification of novel sequences and accessible regions (e.g., DNase I 20 hypersensitive sites) in genes allows for the design of fusion molecules which facilitate regulation of gene expression. Thus, in certain embodiments, the compositions and methods disclosed herein involve fusions between a DNA-binding domain specifically targeted to regulatory regions of a NR gene and a functional (e.g., repression or activation) domain (or a polynucleotide encoding 25 such a fusion). In this way, the repression or activation domain is brought into proximity with a sequence in the NR gene that is bound by the DNA-binding domain. The transcriptional regulatory function of the functional domain is then able to act on NR regulatory sequences.

In additional embodiments, targeted remodeling of chromatin, as disclosed 30 in co-owned U.S. patent application entitled "Targeted Modification of Chromatin Structure," can be used to generate one or more sites in cellular chromatin that are accessible to the binding of a DNA binding molecule.

Fusion molecules are constructed by methods of cloning and biochemical conjugation that are well-known to those of skill in the art. Fusion molecules

comprise a DNA-binding domain and a functional domain (e.g., a transcriptional activation or repression domain). Fusion molecules also optionally comprise nuclear localization signals (such as, for example, that from the SV40 medium T-antigen) and epitope tags (such as, for example, FLAG and hemagglutinin).

5 Fusion proteins (and nucleic acids encoding them) are designed such that the translational reading frame is preserved among the components of the fusion.

Fusions between a polypeptide component of a functional domain (or a functional fragment thereof) on the one hand, and a non-protein DNA-binding domain (e.g., antibiotic, intercalator, minor groove binder, nucleic acid) on the 10 other, are constructed by methods of biochemical conjugation known to those of skill in the art. *See, for example, the Pierce Chemical Company (Rockford, IL) Catalogue.* Methods and compositions for making fusions between a minor groove binder and a polypeptide have been described. Mapp *et al.* (2000) *Proc. Natl. Acad. Sci. USA* 97:3930-3935.

15 The fusion molecules disclosed herein comprise a DNA-binding domain which binds to a target site in a NR gene. In certain embodiments, the target site is present in an accessible region of cellular chromatin. Accessible regions can be determined as described, for example, in co-owned U.S. Patent Application Serial No. 60/228,556. If the target site is not present in an accessible region of cellular 20 chromatin, one or more accessible regions can be generated as described in co-owned U.S. patent application entitled "Targeted Modification of Chromatin Structure." In additional embodiments, the DNA-binding domain of a fusion molecule is capable of binding to cellular chromatin regardless of whether its target site is in an accessible region or not. For example, such DNA-binding 25 domains are capable of binding to linker DNA and/or nucleosomal DNA.

Examples of this type of "pioneer" DNA binding domain are found in certain steroid receptor and in hepatocyte nuclear factor 3 (HNF3). Cordingley *et al.* (1987) *Cell* 48:261-270; Pina *et al.* (1990) *Cell* 60:719-731; and Cirillo *et al.* (1998) *EMBO J.* 17:244-254.

30 Methods of gene regulation targeted to a specific sequence with a DNA binding domain can achieve modulation of gene expression, for example NR gene expression. Modulation of gene expression can be in the form of increased expression or repression. As described herein, repression of NR expression can be used to reduce or prevent tumor formation and/or metastasis and other disease

processes. Alternatively, modulation can be in the form of activation, if activation of gene expression is desired. In this case, cellular chromatin is contacted with a fusion molecule comprising, an activation domain and a DNA-binding domain. Preferably, the DNA-binding domain is specific for a regulatory element of the 5 target gene, e.g., a NR gene.

For such applications, the fusion molecule is typically formulated with a pharmaceutically acceptable carrier, as is known to those of skill in the art. See, for example, Remington's Pharmaceutical Sciences, 17<sup>th</sup> ed., 1985; and co-owned WO 00/42219.

10 The functional component/domain can be selected from any of a variety of different components capable of influencing transcription of a gene once the exogenous molecule binds to an identified regulatory sequence via the DNA binding domain of the exogenous molecule. Hence, the functional component can include, but is not limited to, various transcription factor domains, such as 15 activators, repressors, co-activators, co-repressors, and silencers.

An exemplary functional domain for fusing with a DNA-binding domain such as, for example, a ZFP, to be used for repressing expression of a gene is a KRAB repression domain from the human KOX-1 protein (see, e.g., Thiesen et al., *New Biologist* 2, 363-374 (1990); Margolin et al., *Proc. Natl. Acad. Sci. USA* 91, 4509-4513 (1994); Pemble et al., *Nucl. Acids Res.* 22:2908-2914 (1994); Witzgall et al., *Proc. Natl. Acad. Sci. USA* 91, 4514-4518 (1994). Another suitable repression domain is methyl binding domain protein 2B (MBD-2B) (see, also Hendrich et al. (1999) *Mamm Genome* 10:906-912 for description of MBD proteins). Another useful repression domain is that associated with the v-ErbA 25 protein. See, for example, Damm, et al. (1989) *Nature* 339:593-597; Evans (1989) *Int. J. Cancer Suppl.* 4:26-28; Pain et al. (1990) *New Biol.* 2:284-294; Sap et al. (1989) *Nature* 340:242-244; Zenke et al. (1988) *Cell* 52:107-119; and Zenke et al. (1990) *Cell* 61:1035-1049.

Suitable domains for achieving activation include the HSV VP16 30 activation domain (see, e.g., Hagmann et al., *J. Virol.* 71, 5952-5962 (1997)) nuclear hormone receptors (see, e.g., Torchia et al., *Curr. Opin. Cell. Biol.* 10:373-383 (1998)); the p65 subunit of nuclear factor kappa B (Bitko & Barik, *J. Virol.* 72:5610-5618 (1998) and Doyle & Hunt, *Neuroreport* 8:2937-2942 (1997)); Liu et

al., *Cancer Gene Ther.* 5:3-28 (1998)), or artificial chimeric functional domains such as VP64 (Seifpal et al., *EMBO J.* 11, 4961-4968 (1992)).

Additional exemplary activation domains include, but are not limited to, VP16, VP64, p300, CBP, PCAF, SRC1 PvALF, AtHD2A and ERF-2. *See, for example, Robyr et al. (2000) Mol. Endocrinol.* 14:329-347; *Collingwood et al. (1999) J. Mol. Endocrinol.* 23:255-275; *Leo et al. (2000) Gene* 245:1-11; *Manteuffel-Cymborowska (1999) Acta Biochim. Pol.* 46:77-89; *McKenna et al. (1999) J. Steroid Biochem. Mol. Biol.* 69:3-12; *Malik et al. (2000) Trends Biochem. Sci.* 25:277-283; and *Lemon et al. (1999) Curr. Opin. Genet. Dev.* 9:499-504. Additional exemplary activation domains include, but are not limited to, OsGAI, HALF-1, C1, AP1, ARF-5, -6, -7, and -8, CPRF1, CPRF4, MYC-RP/GP, and TRAB1. *See, for example, Ogawa et al. (2000) Gene* 245:21-29; *Okanami et al. (1996) Genes Cells* 1:87-99; *Goff et al. (1991) Genes Dev.* 5:298-309; *Cho et al. (1999) Plant Mol. Biol.* 40:419-429; *Ulmason et al. (1999) Proc. Natl. Acad. Sci. USA* 96:5844-5849; *Sprenger-Haussels et al. (2000) Plant J.* 22:1-8; *Gong et al. (1999) Plant Mol. Biol.* 41:33-44; and *Hobo et al. (1999) Proc. Natl. Acad. Sci. USA* 96:15,348-15,353.

Additional exemplary repression domains include, but are not limited to, KRAB, SID, MBD2, MBD3, members of the DNMT family (e.g., DNMT1, DNMT3A, DNMT3B), Rb, and MeCP2. *See, for example, Bird et al. (1999) Cell* 99:451-454; *Tyler et al. (1999) Cell* 99:443-446; *Knoepfler et al. (1999) Cell* 99:447-450; and *Robertson et al. (2000) Nature Genet.* 25:338-342. Additional exemplary repression domains include, but are not limited to, ROM2 and AtHD2A. *See, for example, Chern et al. (1996) Plant Cell* 8:305-321; and *Wu et al. (2000) Plant J.* 22:19-27.

Additional functional domains are disclosed, for example, in co-owned WO 00/41566.

#### **Polynucleotide and Polypeptide Delivery**

The compositions described herein can be provided to the target cell *in vitro* or *in vivo*. In addition, the compositions can be provided as polypeptides, polynucleotides or combination thereof.

##### **A. Delivery of Polynucleotides**

In certain embodiments, the compositions are provided as one or more polynucleotides. Further, as noted above, the compositions described herein may

be designed as a fusion between a DNA-binding domain targeted to a gene (e.g., NR gene) and a functional domain (e.g., repressive domain) and can be encoded by a fusion nucleic acid. In both fusion and non-fusion cases, the nucleic acid can be cloned into intermediate vectors for transformation into prokaryotic or 5 eukaryotic cells for replication and/or expression. Intermediate vectors for storage or manipulation of the nucleic acid or production of protein can be prokaryotic vectors, (e.g., plasmids), shuttle vectors, insect vectors, or viral vectors for example. A nucleic acid can also be cloned into an expression vector, for administration to a bacterial cell, fungal cell, protozoal cell, plant cell, or animal 10 cell, preferably a mammalian cell, more preferably a human cell.

To obtain expression of a cloned nucleic acid, it is typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook *et al.*, *supra*; Ausubel *et al.*, *supra*; and Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990). Bacterial expression systems are 15 available in, e.g., *E. coli*, *Bacillus sp.*, and *Salmonella*. Palva *et al.* (1983) *Gene* 22:229-235. Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available, for example, from 20 Invitrogen, Carlsbad, CA and Clontech, Palo Alto, CA.

The promoter used to direct expression of the nucleic acid of choice depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification. In contrast, when a protein is to be used *in vivo*, either a constitutive or an inducible promoter is used, depending 25 on the particular use of the protein. In addition, a weak promoter can be used, such as HSV TK or a promoter having similar activity. The promoter typically can also include elements that are responsive to transactivation, e.g., hypoxia response elements, Gal4 response elements, *lac* repressor response element, and small molecule control systems such as tet-regulated systems and the RU-486 system. See, e.g., Gossen *et al.* (1992) *Proc. Natl. Acad. Sci USA* 89:5547-5551; Oligino *et al.* (1998) *Gene Ther.* 5:491-496; Wang *et al.* (1997) *Gene Ther.* 30 4:432-441; Neering *et al.* (1996) *Blood* 88:1147-1155; and Rendahl *et al.* (1998) *Nat. Biotechnol.* 16:757-761.

In addition to a promoter, an expression vector typically contains a transcription unit or expression cassette that contains additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, *e.g.*, to the nucleic acid sequence, and signals required, *e.g.*, for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding, and/or translation termination. Additional elements of the cassette may include, *e.g.*, enhancers, and heterologous spliced intronic signals.

The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the resulting polypeptide, *e.g.*, expression in plants, animals, bacteria, fungi, protozoa *etc.* Standard bacterial expression vectors include plasmids such as pBR322, pBR322-based plasmids, pSKF, pET23D, and commercially available fusion expression systems such as GST and LacZ. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, *e.g.*, c-myc or FLAG.

Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, *e.g.*, SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High-yield expression systems are also suitable, such as baculovirus vectors in insect cells, for example under the transcriptional control of the polyhedrin promoter or any other strong baculovirus promoter.

Elements that are typically included in expression vectors also include a replicon that functions in *E. coli* (or in the prokaryotic host, if other than *E. coli*), a selective marker, *e.g.*, a gene encoding antibiotic resistance, to permit selection of

bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the vector to allow insertion of recombinant sequences.

Standard transfection methods can be used to produce bacterial, mammalian, yeast, insect, or other cell lines that express large quantities of 5 proteins, which can be purified, if desired, using standard techniques. *See, e.g.,* Colley *et al.* (1989) *J. Biol. Chem.* 264:17619-17622; and *Guide to Protein Purification, in Methods in Enzymology*, vol. 182 (Deutscher, ed.) 1990. Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques. *See, e.g.,* Morrison (1977) *J. Bacteriol.* 132:349-351; Clark-10 Curtiss *et al.* (1983) in *Methods in Enzymology* 101:347-362 (Wu *et al.*, eds).

Any procedure for introducing foreign nucleotide sequences into host cells can be used. These include, but are not limited to, the use of calcium phosphate transfection, DEAE-dextran-mediated transfection, polybrene, protoplast fusion, 15 electroporation, lipid-mediated delivery (*e.g.*, liposomes), microinjection, particle bombardment, introduction of naked DNA, plasmid vectors, viral vectors (both episomal and integrative) and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.,* Sambrook *et al., supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully 20 introducing at least one gene into the host cell capable of expressing the protein of choice.

Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids into mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding reprogramming polypeptides to 25 cells *in vitro*. Preferably, nucleic acids are administered for *in vivo* or *ex vivo* gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For reviews 30 of gene therapy procedures, *see, for example*, Anderson (1992) *Science* 256:808-813; Nabel *et al.* (1993) *Trends Biotechnol.* 11:211-217; Mitani *et al.* (1993) *Trends Biotechnol.* 11:162-166; Dillon (1993) *Trends Biotechnol.* 11:167-175; Miller (1992) *Nature* 357:455-460; Van Brunt (1988) *Biotechnology* 6(10):1149-1154; Vigne (1995) *Restorative Neurology and Neuroscience* 8:35-36; Kremer *et*

*al.* (1995) *British Medical Bulletin* 51(1):31-44; Haddada *et al.*, in *Current Topics in Microbiology and Immunology*, Doerfler and Böhm (eds), 1995; and Yu *et al.* (1994) *Gene Therapy* 1:13-26.

Methods of non-viral delivery of nucleic acids include lipofection, 5 microinjection, ballistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in, *e.g.*, U.S. Patent Nos. 5,049,386; 4,946,787; and 4,897,355 and lipofection reagents are sold commercially (*e.g.*, Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for 10 efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424 and WO 91/16024. Nucleic acid can be delivered to cells (*ex vivo* administration) or to target tissues (*in vivo* administration).

The preparation of lipid:nucleic acid complexes, including targeted 15 liposomes such as immunolipid complexes, is well known to those of skill in the art. *See, e.g.*, Crystal (1995) *Science* 270:404-410; Blaese *et al.* (1995) *Cancer Gene Ther.* 2:291-297; Behr *et al.* (1994) *Bioconjugate Chem.* 5:382-389; Remy *et al.* (1994) *Bioconjugate Chem.* 5:647-654; Gao *et al.* (1995) *Gene Therapy* 2:710-722; Ahmad *et al.* (1992) *Cancer Res.* 52:4817-4820; and U.S. Patent 20 Nos. 4,186,183; 4,217,344; 4,235,871; 4,261,975; 4,485,054; 4,501,728; 4,774,085; 4,837,028 and 4,946,787.

The use of RNA or DNA virus-based systems for the delivery of nucleic acids take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in* 25 *vitro*, wherein the modified cells are administered to patients (*ex vivo*).

Conventional viral based systems for the delivery of ZFPs include retroviral, lentiviral, poxviral, adenoviral, adeno-associated viral, vesicular stomatitis viral and herpesviral vectors. Integration in the host genome is possible with certain 30 viral vectors, including the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, allowing alteration and/or expansion of the potential target cell population. Lentiviral vectors are retroviral vector that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a 5 retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors have a packaging capacity of up to 6-10 kb of foreign sequence and are comprised of *cis*-acting long terminal repeats (LTRs). The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide 10 permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), simian immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof. Buchscher *et al.* (1992) *J. Virol.* 66:2731-2739; Johann *et al.* (1992) *J. Virol.* 66:1635-1640; Sommerfelt *et al.* (1990) *Virol.* 176:58-59; 15 Wilson *et al.* (1989) *J. Virol.* 63:2374-2378; Miller *et al.* (1991) *J. Virol.* 65:2220-2224; and PCT/US94/05700).

Adeno-associated virus (AAV) vectors are also used to transduce cells with target nucleic acids, *e.g.*, in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures. *See, e.g.*, West *et al.* (1987) *Virology* 160:38-47; U.S. Patent No. 4,797,368; WO 93/24641; Kotin 20 (1994) *Hum. Gene Ther.* 5:793-801; and Muzyczka (1994) *J. Clin. Invest.* 94:1351. Construction of recombinant AAV vectors are described in a number of publications, including U.S. Patent No. 5,173,414; Tratschin *et al.* (1985) *Mol. Cell. Biol.* 5:3251-3260; Tratschin, *et al.* (1984) *Mol. Cell. Biol.* 4:2072-2081; 25 Hermonat *et al.* (1984) *Proc. Natl. Acad. Sci. USA* 81:6466-6470; and Samulski *et al.* (1989) *J. Virol.* 63:3822-3828.

Recombinant adeno-associated virus vectors based on the defective and nonpathogenic parvovirus adeno-associated virus type 2 (AAV-2) are a promising gene delivery system. Exemplary AAV vectors are derived from a plasmid 30 containing the AAV 145 bp inverted terminal repeats flanking a transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. Wagner *et al.* (1998) *Lancet* 351 (9117):1702-3; and Kearns *et al.* (1996) *Gene Ther.* 9:748-55.

pLASN and MFG-S are examples are retroviral vectors that have been used in clinical trials. Dunbar *et al.* (1995) *Blood* 85:3048-305; Kohn *et al.* (1995) *Nature Med.* 1:1017-102; Malech *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:12133-12138. PA317/pLASN was the first therapeutic vector used in a gene 5 therapy trial. Blaese *et al.* (1995) *Science* 270:475-480. Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. Ellem *et al.* (1997) *Immunol Immunother.* 44(1):10-20; Dranoff *et al.* (1997) *Hum. Gene Ther.* 1:111-2.

In applications for which transient expression is preferred, adenoviral-10 based systems are useful. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and are capable of infecting, and hence delivering nucleic acid to, both dividing and non-dividing cells. With such vectors, high titers and levels of expression have been obtained. Adenovirus vectors can be produced in large quantities in a relatively simple system.

15 Replication-deficient recombinant adenovirus (Ad) vectors can be produced at high titer and they readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and/or E3 genes; the replication defector vector is propagated in human 293 cells that supply the required E1 functions *in trans*. Ad vectors can transduce 20 multiple types of tissues *in vivo*, including non-dividing, differentiated cells such as those found in the liver, kidney and muscle. Conventional Ad vectors have a large carrying capacity for inserted DNA. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for antitumor immunization with intramuscular injection. Sterman *et al.* (1998) *Hum. Gene Ther.* 7:1083-1089.

25 Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenzweig *et al.* (1996) *Infection* 24:5-10; Sterman *et al.*, *supra*; Welsh *et al.* (1995) *Hum. Gene Ther.* 2:205-218; Alvarez *et al.* (1997) *Hum. Gene Ther.* 5:597-613; and Topf *et al.* (1998) *Gene Ther.* 5:507-513.

30 Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and  $\Psi$ 2 cells or PA317 cells, which package retroviruses. Viral vectors used in gene therapy are usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral

sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the protein to be expressed. Missing viral functions are supplied in *trans*, if necessary, by the packaging cell line. For example, AAV vectors used in gene therapy typically 5 only possess ITR sequences from the AAV genome, which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and 10 expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment, which preferentially inactivates adenoviruses.

In many gene therapy applications, it is desirable that the gene therapy 15 vector be delivered with a high degree of specificity to a particular tissue type. A viral vector can be modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the outer surface of the virus. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han *et al.* (1995) *Proc. Natl. Acad. Sci. USA* 92:9747-9751 reported that Moloney murine leukemia virus can be modified 20 to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other pairs of virus expressing a ligand fusion protein and target cell expressing a receptor. For example, 25 filamentous phage can be engineered to display antibody fragments (*e.g.*, F<sub>ab</sub> or F<sub>v</sub>) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to non-viral vectors. Such vectors can be engineered to contain specific uptake sequences thought to favor uptake by specific target cells.

30 Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (*e.g.*, intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described *infra*. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (*e.g.*, lymphocytes, bone

marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

Ex vivo cell transfection for diagnostics, research, or for gene therapy (e.g., 5 via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with a nucleic acid (gene or cDNA), and re-infused back into the subject organism (e.g., patient). Various cell types suitable for ex vivo transfection are well known to those of skill in the art. See, e.g., Freshney *et 10 al.*, *Culture of Animal Cells, A Manual of Basic Technique*, 3rd ed., 1994, and references cited therein, for a discussion of isolation and culture of cells from patients.

In one embodiment, hematopoietic stem cells are used in ex vivo 15 procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types *in vitro*, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ stem cells *in vitro* into clinically important immune cell types using cytokines such a GM-CSF, IFN- $\gamma$  and TNF- $\alpha$  are known. Inaba *et al.* (1992) *J. Exp. Med.* 176:1693-1702.

20 Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells). See Inaba *et al.*, *supra*.

25 Vectors (e.g., retroviruses, adenoviruses, liposomes, *etc.*) containing therapeutic nucleic acids can be also administered directly to the organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. Suitable methods of administering 30 such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions, as described below. See, 5 *e.g., Remington's Pharmaceutical Sciences*, 17th ed., 1989.

#### B. Delivery of Polypeptides

In other embodiments, for example in certain *in vitro* situations, the target 10 cells are cultured in a medium containing a functional domain (or functional fragments thereof) fused to a targeted (e.g., NR-targeted) DNA binding domain.

An important factor in the administration of polypeptide compounds is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. 15 Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins, lipids and other compounds, which have the ability to translocate polypeptides across a cell membrane, have been described.

For example, "membrane translocation polypeptides" have amphiphilic or 20 hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, *Antennapedia*, was found to be the third helix of the protein, from amino acid position 43 to 58. Prochiantz (1996) *Curr. Opin. Neurobiol.* 6:629-634. Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation 25 characteristics. Lin *et al.* (1995) *J. Biol. Chem.* 270:14255-14258.

Examples of peptide sequences which can be linked to a NR-targeted functional polypeptide for facilitating its uptake into cells include, but are not 30 limited to: an 11 amino acid peptide of the *tat* protein of HIV; a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (*see* Fahraeus *et al.* (1996) *Curr. Biol.* 6:84); the third helix of the 60-amino acid long homeodomain of *Antennapedia* (Derossi *et al.* (1994) *J. Biol. Chem.* 269:10444); the h region of a signal peptide, such as the Kaposi fibroblast growth factor (K-

FGF) h region (Lin *et al.*, *supra*); and the VP22 translocation domain from HSV (Elliot *et al.* (1997) *Cell* 88:223-233). Other suitable chemical moieties that provide enhanced cellular uptake can also be linked, either covalently or non-covalently, to the polypeptides described herein.

5       Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules (called "binary toxins") are composed of at least two parts: a translocation or binding domain and a separate toxin domain. Typically, the translocation domain, which can optionally be a polypeptide, binds to a cellular receptor, facilitating transport of the toxin into the cell. Several 10 bacterial toxins, including *Clostridium perfringens* iota toxin, diphtheria toxin (DT), *Pseudomonas* exotoxin A (PE), pertussis toxin (PT), *Bacillus anthracis* toxin, and pertussis adenylate cyclase (CYA), have been used to deliver peptides to the cell cytosol as internal or amino-terminal fusions. Arora *et al.* (1993) *J. Biol. Chem.* 268:3334-3341; Perelle *et al.* (1993) *Infect. Immun.* 61:5147-5156; 15 Stenmark *et al.* (1991) *J. Cell Biol.* 113:1025-1032; Donnelly *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:3530-3534; Carbonetti *et al.* (1995) *Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295; Sebo *et al.* (1995) *Infect. Immun.* 63:3851-3857; Klimpel *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10277-10281; and Novak *et al.* (1992) *J. Biol. Chem.* 267:17186-17193.

20       Such subsequences can be used to translocate polypeptides, including the polypeptides as disclosed herein, across a cell membrane. This is accomplished, for example, by derivatizing the fusion polypeptide with one of these translocation sequences, or by forming an additional fusion of the translocation sequence with the fusion polypeptide. Optionally, a linker can be used to link the fusion 25 polypeptide and the translocation sequence. Any suitable linker can be used, e.g., a peptide linker.

A suitable polypeptide can also be introduced into an animal cell, preferably a mammalian cell, via liposomes and liposome derivatives such as immunoliposomes. The term "liposome" refers to vesicles comprised of one or 30 more concentrically ordered lipid bilayers, which encapsulate an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell.

The liposome fuses with the plasma membrane, thereby releasing the compound into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the

liposome is either degraded or it fuses with the membrane of the transport vesicle and releases its contents.

In current methods of drug delivery *via* liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer is degraded over time through the action of various agents in the body. Alternatively, active drug release involves using an agent to induce a permeability change in the liposome vesicle. Liposome membranes can be constructed so that they become destabilized when the environment becomes acidic near the liposome membrane.

See, e.g., *Proc. Natl. Acad. Sci. USA* 84:7851 (1987); *Biochemistry* 28:908 (1989). When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents. This destabilization is termed fusogenesis. Dioleoylphosphatidylethanolamine (DOPE) is the basis of many "fusogenic" systems.

For use with the methods and compositions disclosed herein, liposomes typically comprise a fusion polypeptide as disclosed herein, a lipid component, e.g., a neutral and/or cationic lipid, and optionally include a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (e.g., an antigen). A variety of methods are available for preparing liposomes as described in, e.g.; U.S. Patent Nos. 4,186,183; 4,217,344; 4,235,871; 4,261,975; 4,485,054; 4,501,728; 4,774,085; 4,837,028; 4,235,871; 4,261,975; 4,485,054; 4,501,728; 4,774,085; 4,837,028; 4,946,787; PCT Publication No. WO 91/17424; Szoka *et al.* (1980) *Ann. Rev. Biophys. Bioeng.* 9:467; Deamer *et al.* (1976) *Biochim. Biophys. Acta* 443:629-634; Fraley, *et al.* (1979) *Proc. Natl. Acad. Sci. USA* 76:3348-3352; Hope *et al.* (1985) *Biochim. Biophys. Acta* 812:55-65; Mayer *et al.* (1986) *Biochim. Biophys. Acta* 858:161-168; Williams *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:242-246; *Liposomes*, Ostro (ed.), 1983, Chapter 1); Hope *et al.* (1986) *Chem. Phys. Lip.* 40:89; 30 Gregoriadis, *Liposome Technology* (1984) and Lasic, *Liposomes: from Physics to Applications* (1993). Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are well known in the art.

In certain embodiments, it may be desirable to target a liposome using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (e.g., ligands, receptors, and monoclonal antibodies) has been previously described. *See, e.g.,* 5 U.S. Patent Nos. 4,957,773 and 4,603,044.

Examples of targeting moieties include monoclonal antibodies specific to antigens associated with neoplasms, such as prostate cancer specific antigen and MAGE. Tumors can also be diagnosed by detecting gene products resulting from the activation or over-expression of oncogenes, such as *ras* or *c-erbB2*. In 10 addition, many tumors express antigens normally expressed by fetal tissue, such as the alphafetoprotein (AFP) and carcinoembryonic antigen (CEA). Sites of viral infection can be diagnosed using various viral antigens such as hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV-1) and papilloma virus 15 antigens. Inflammation can be detected using molecules specifically recognized by surface molecules which are expressed at sites of inflammation such as integrins (e.g., VCAM-1), selectin receptors (e.g., ELAM-1) and the like.

Standard methods for coupling targeting agents to liposomes are used. These methods generally involve the incorporation into liposomes of lipid 20 components, *e.g.*, phosphatidylethanolamine, which can be activated for attachment of targeting agents, or incorporation of derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A. *See* Renneisen *et al.* (1990) *J. Biol. Chem.* 265:16337-16342 and Leonetti *et al.* 25 (1990) *Proc. Natl. Acad. Sci. USA* 87:2448-2451.

#### Pharmaceutical compositions and administration

Targeted DNA binding domains (*e.g.*, a zinc finger protein (ZFP)) and functional domains as disclosed herein, and expression vectors encoding these 30 polypeptides, can be used in conjunction with various methods of gene therapy to facilitate the action of a therapeutic gene product. In such applications, the ZFP can be administered directly to a patient to facilitate the modulation of gene expression and for therapeutic or prophylactic applications, for example, cancer, ischemia, diabetic retinopathy, macular degeneration, rheumatoid arthritis,

psoriasis, HIV infection, sickle cell anemia, Alzheimer's disease, muscular dystrophy, neurodegenerative diseases, vascular disease, cardiovascular disease, cystic fibrosis, stroke, and the like. Examples of microorganisms whose replication and/or pathogenicity can be inhibited through use of the methods and compositions disclosed herein include pathogenic bacteria, *e.g.*, Chlamydia, Rickettsial bacteria, Mycobacteria, Staphylococci, Streptococci, Pneumococci, Meningococci and Conococci, Klebsiella, Proteus, Serratia, Pseudomonas, Legionella, Diphtheria, Salmonella, Bacilli (*e.g.*, anthrax), Vibrio (*e.g.*, cholera), Clostridium (*e.g.*, tetanus, botulism), Yersinia (*e.g.*, plague), Leptospirosis, and Borrellia (*e.g.*, Lyme disease bacteria); infectious fungus, *e.g.*, *Aspergillus*, *Candida* species; protozoa such as sporozoa (*e.g.*, *Plasmodia*), rhizopods (*e.g.*, *Entamoeba*) and flagellates (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, *etc.*); viruses, *e.g.*, hepatitis (A, B, or C), herpes viruses (*e.g.*, VZV, HSV-1, HHV-6, HSV-II, CMV, and EBV), HIV, Ebola, Marburg and related hemorrhagic fever-causing viruses, adenoviruses, influenza viruses, flaviviruses, echoviruses, rhinoviruses, coxsackie viruses, coronaviruses, respiratory syncytial viruses, mumps viruses, rotaviruses, measles viruses, rubella viruses, parvoviruses, vaccinia viruses, HTLV viruses, retroviruses, lentiviruses, dengue viruses, papillomaviruses, polioviruses, rabies viruses, and arboviral encephalitis viruses, *etc.*

Administration of therapeutically effective amounts of regulatory polypeptides or nucleic acids encoding these fusion polypeptides is by any of the routes normally used for introducing polypeptides or nucleic acids into ultimate contact with the tissue to be treated. The polypeptides or nucleic acids are administered in any suitable manner, preferably with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers or excipients are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions. *See, e.g.*, *Remington's Pharmaceutical Sciences*, 17<sup>th</sup> ed. 1985.

5 Polypeptides or nucleic acids, alone or in combination with other suitable components, can be made into aerosol formulations (*i.e.*, they can be “nebulized”) to be administered *via* inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

10 Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. Compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or 15 multi-dose sealed containers, such as ampoules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind known to those of skill in the art.

#### Advantages

20 The compositions and methods disclosed herein will enhance and improve upon existing methods of analyzing various genes. Mouse knockout models, for example, require researchers to breed and isolate animals that are homozygous for the disabled gene. Using a targeted transcriptional repressor, as disclosed herein, relieves researchers of this need, since it will have a *trans*-dominant effect inside 25 cells. Furthermore, some knockouts can be lethal. To overcome this complication, repressor-ZFPs are integrated into the genome as inducible genes that can be switched on (at an appropriate time in development) to repress a gene's expression. Although other technologies for repression are available (principally antisense technology and ribozymes), ZFP-based transcription factors 30 have several features that make them an attractive alternative. First, artificial transcription factors mimic the natural processes of gene regulation. Exploiting the natural components of cells in this way also offers the opportunity to combine a designed ZFP domain with naturally occurring effector domains such as transcriptional activators and repressors. Second, targeting genes, rather than

RNA or proteins, has the advantage of affecting regulation prior to gene expression. There are usually only two copies of each gene per cell, compared to the 100-10,000 copies of each mRNA and even more copies of each protein. Hence, antisense repression of genes can be an inefficient process, and although 5 ribozymes can potentially catalyze the degradation of many transcripts, recent studies have shown that the relatively low intracellular MgCl<sub>2</sub> concentrations greatly reduces their activity (88). Finally, ZFP-based transcription factors can be used to activate or repress genes making them an extremely flexible platform.

The following examples are presented as illustrative of, but not limiting, 10 the claimed subject matter. All patents, patent applications and publications mentioned herein are hereby incorporated by reference, in their entireties.

### EXAMPLES

#### **Example 1: Characterization of the Promoter Region of the**

##### **15 Endogenous Estrogen Receptor- $\alpha$ Gene**

The promoter regions of human ER- $\alpha$  gene (this example) and the mouse PPAR- $\gamma$  gene (see Example 2) have been mapped. For the human ER- $\alpha$ , the promoter region was determined from existing sequences from Genbank. Assembly of the promoter region involved connecting 3 overlapping fragments 20 (GI: 3550293, 35159, and 4503602). An additional kilobase of 5'-sequence was extrapolated from the working clone of 6q25.1 (accession no. RP11-237E17). Proper alignment of the sequence was confirmed with a PCR screen using one primer designed from the 5' fragment and another designed from an adjacent fragment.

25 Using the assembled sequence, a probe was designed to an Xba I digested genomic fragment that had been treated with increasing concentrations of DNase I prior to restriction digestion. See co-owned U.S. Patent Application Serial No. 60/228,556 for details of the DNase digestion and indirect end-labeling techniques. Briefly, separate aliquots of permeabilized cells were exposed to 30 different concentrations of DNaseI; DNA was extracted, digested with a restriction enzyme, the fragments were separated on an agarose gel, and the gel was blotted. The blot was then probed with a labeled fragment which was located within the vicinity of the gene of interest, one of whose ends was defined by the

restriction enzyme used for digestion of the DNA. Comparison of two breast cancer cell lines elucidated a hypersensitive site in the region of the P<sub>1</sub> transcriptional start site in the MCF-7 cell line (ER positive). Interestingly, no corresponding hypersensitive site was detected in the MDA-MB-231 (ER negative) cell line. See Figure 2.

Using the information derived from hypersensitive site analysis, three-finger ZFPs were designed to target sequences within the hypersensitive region. The target nucleotide sequences, and the amino acid sequences of the recognition helices of the zinc fingers used to target these sequences, are listed in Table 2. 10 (See, also, Figure 7). The binding affinity of five of these proteins has been experimentally determined using quantitative electrophoretic gel mobility shift assay (EMSA; see experimental design and methods). These ZFPs are linked to activation or repression domains to modulate expression of the ER- $\alpha$  locus.

Position from P <sub>1</sub>	Strand	Sequence	ZFP1	ZFP2	ZFP3	Kd (nM)
-552	Minus	GGGGCGGAG	*RSDNLTR	*RSDELQR	*RSDHLSR	0.400
-514	Minus	GCAGCTGGG	*RSDHLAR	*QSSDLTR	*QSGDLTR	0.019
-294	Minus	GGGGCCGGC	*DRSHLTR	*ERGTLAR	*RSDHLSR	0.025
-258	Minus	GACTGGGCT	*QSSDLTR	*RSDDLTK	*DRSNLTR	U
-191	Minus	GAGGCTGAG	*RSDNLTR	*QSSDLQR	*RSDNLVR	0.020
-158	Plus	AGGGAAAGCT	*QSSDLTR	*QSGNLAR	*RSDHLTQ	U
3	Plus	GGAGCTGGC	*DRSHLTR	*QSSDLSR	*QSGHLQR	U
214	Minus	GGTGCAGAC	*DRSNLTR	*QSGDLTR	*MSHHLSR	0.100
509	Minus	GTTGGAGCC	*ERGTLAR	*QSGHLQR	*QSSALTR	U

15 Table 2. List of engineered zinc finger proteins. ZFPs were selected to recognize sequences around the hypersensitive site (-50), which is in the vicinity of the transcription start site (P<sub>1</sub>). DNA target sequence is indicated as well as the seven amino acids comprising the recognition helix of each zinc finger. The Kd was determined by gel mobility shift assays (see experimental design and methods).

20

#### Example 2: Repression of the Nuclear Receptor PPAR $\gamma$ Gene

PPAR- $\gamma$  is a nuclear receptor that is intimately involved in the transcriptional control of adipogenesis (13). In the process of adipogenesis, a shift in the gene expression profile occurs. This is due in part, to the expression of 25 PPAR- $\gamma$ , which controls the expression of several fat-cell specific genes (e.g., aP2, PEPCK, and LEPTIN) (13). There are two isoforms of PPAR- $\gamma$ , which arise from alternative promoter usage. PPAR- $\gamma$ 2 is 30 amino acids longer than PPAR- $\gamma$ 1 and is the predominant form expressed in the adipocyte (81). This example

demonstrates repression of the expression of the endogenous PPAR- $\gamma$  gene.

Successful repression of PPAR- $\gamma$ 2 was achieved by first characterizing the promoter region. Two DNase I hypersensitive sites (HS1 and HS2) were identified within the promoter region of PPAR- $\gamma$ 2 (Figure 3).

5 For repression, HS1 was targeted because of its proximity to the transcriptional start site. Three ZFPs (zfp52, zfp54 and zfp55) were designed to recognize target sites in a region 200-300 base pairs upstream of the transcription start. A retrovirus containing either zfp54 or 55 fused to the KRAB domain was transduced into 3T3-L1 fibroblast cells. Infection efficiency was near 100%. As  
10 a control, a retrovirus containing the *lacZ* gene was also transduced to determine if transduction had an effect on the cells. PPAR $\gamma$  mRNA levels were measured by the TaqMan<sup>®</sup> real-time PCR procedure (Roche, Indianapolis, IN). Induction of PPAR- $\gamma$  occurs upon stimulation of adipogenesis. Prior to adipogenesis, relatively low levels of both PPAR- $\gamma$ s are expressed (Figure 4). During adipogenesis, there  
15 is an increase in gene expression of both PPAR- $\gamma$ 1,2. However, in the presence of either KRAB-zfp54 or KRAB-zfp55, induction of PPAR- $\gamma$ 2 from promoter B was inhibited. Importantly, transcription from the upstream promoter (Promoter A; PPAR- $\gamma$ 1) was uninhibited by the presence of the downstream KRAB-ZFPs. This result indicates that both zfp54 and zfp55 were specific for repression of promoter  
20 B, demonstrating the ability to selectively repress one isoform of a gene without affecting the other. In summary, these results demonstrate the feasibility of specific targeting of regulatory molecules to specific promoter elements within a gene.

25 **Example 3: Identification of promoter regions of genes encoding nuclear hormone receptors**

To efficiently regulate gene expression of target nuclear receptors, the regulatory elements within and upstream of the promoter need to be defined. This will allow for the systematic identification of potential target sequences. As a first  
30 step, genomic sequence for *Homo sapiens* and *Mus musculus* are obtained from the available sequences. Available sequences are extended by Genomic 'Walking,' if necessary. The entire promoter region of ER- $\alpha$ , ER- $\beta$ , and AR from *Homo Sapiens* has been extrapolated from the public databases at the National

Center for Biotechnology Information (NCBI). Briefly, cDNA sequences were placed under an advance Basic Local Alignment Search Tool (BLAST) search to pull up available genomic sequence. Alignment sequences with high scores were more thoroughly analyzed to determine if the promoter region could be obtained.

5 In many instances, the promoter region was determined by connecting overlapping clones. Correctly assembled promoter regions were evaluated by designing primers from overlapping clones to yield a PCR product of defined length. A correctly assembled promoter region yields a PCR product of the predicted size.

10 For target genes in which sufficient sequence is not available, genomic walking techniques (Figure 5A) are employed. In one embodiment, the GenomeWalker kit (Clontech, Palo Alto, CA), is used to determine upstream and downstream sequences surrounding the promoter. Briefly, the procedure involves using four uncloned, adaptor-ligated genomic fragment 'libraries' as templates for 15 gene amplification. A primary PCR amplification contains an outer adaptor primer (AP1, provided in the kit) and a gene-specific primer (GSP1) that is designed based on available sequence. The primary amplification is then used as the template for a secondary round of PCR using another set of adaptor primer (AP2) and a second, nested gene-specific primer (GSP2). This generally results in 20 the one major product from each library. Each DNA fragment has a known 5' end, based on the second gene-specific primer, and can be cloned for sequence analysis.

Another important requirement in the identification of the promoter sequence is to identify the transcriptional start site. This is important in narrowing 25 down the regions of interest for design of ZFPs. Using the RLM-RACE kit (Ambion, Austin, TX) the 5'-untranslated region (UTR) and the transcriptional start site (Figure 5B) are determined. This procedure enriches the isolation of the entire 5'-UTR by selecting only intact (capped) mRNA. Briefly, isolated RNA is treated with calf intestinal phosphatase (CIP) to remove the free 5'-phosphate 30 from rRNA, tRNA, partial (uncapped) mRNA, and contaminating genomic DNA. The sample is treated with tobacco acid pyrophosphatase (TAP) to remove the cap structure from intact mRNA, allowing the direct ligation of a RNA adapter to the 5' end. Ligation is limited to intact mRNA since it still retains a 5'-phosphate. RT-PCR is then performed, leading to the production and amplification of a

cDNA with the full-length 5' UTR. Comparison of a mRNA sequence obtained by this method, with genomic sequences obtained as described above, allows mapping of the transcriptional startsite for a gene of interest. Accordingly, it is possible to systematically characterize the promoter regions of various NRs.

5

**Example 4: Regulation of the ER- $\alpha$  gene by ZFPs**

A number of the ZFP DNA-binding domains were fused to functional domains and tested for their ability to regulate expression of the ER in living cells. The functional domains used in these experiments were the VP16 and p65 activation domains, and the cells used were the MC human prostate cancer cell (MCF-7).

Nucleic acid vectors encoding fusion molecules comprising a given ZFP DNA-binding domain, a VP16 or p65 activation domain, a nuclear localization signal and an epitope tag were constructed as described, for example in co-owned 15 WO 00/41566 and WO 00/42219, Zhang *et al.* (2000) *J. Biol. Chem.* 275:33,850-33,860 and Liu *et al.* (2001) *J. Biol. Chem.* 276:11,323-11,334, the disclosures of which are hereby incorporated by reference in their entireties. Cells were cultured and transfected as described, for example in co-owned WO 00/41566 and WO 00/42219, Zhang *et al.* (2000) *J. Biol. Chem.* 275:33,850-33,860 and Liu *et al.* 20 (2001) *J. Biol. Chem.* 276:11,323-11,334, the disclosures of which are hereby incorporated by reference in their entireties. Total RNA was either isolated from cultured cells using a RNeasy mini-prep kit (Qiagen, Valencia, CA) or purchased from Clontech (Palo Alto, CA). A relative quantitation with standard curve method (Applied BioSystems, Foster City, CA, TaqMan User Bulletin #2) was 25 used to quantitate mRNA levels in each RNA preparation. NKF or NVF were used to normalize the total RNA input for each reaction. Results are shown in Figures 6 and 8.

30 **Example 5: Characterization of the promoter regions of nuclear receptors and identification of putative upstream regulatory elements**

*Mapping of DNase I Hypersensitive Regions.* Using a promoter sequence determined as described in Example 3, probes are designed to map DNase I

accessibility within that region. DNase I nicks each strand of DNA at ~10 bp intervals in the presence of Mg<sup>2+</sup> and Ca<sup>2+</sup>. Briefly, the procedure involves permeabilization of cells using the mild, nonionic detergent IGEPAL (0.5% final) in the presence of 5 mM divalent and 70 mM monovalent cations, (concentrations 5 similar to those present *in vivo*). Aliquots of permeabilized cells are exposed to different concentrations of DNase I, which readily diffuses into the permeabilized cells and into the nucleus. DNase I cleaves cellular chromatin in regions that are more accessible, thus defining one or more regions of sensitivity. This general sensitivity implies a region that is decondensed or 'open'. In most cases, active 10 regulatory elements exhibit DNase I hypersensitivity in the context of chromatin. This approach is minimally invasive and minimally disruptive of nuclear architecture, allowing analysis of chromatin structure in its native state.

15 Transcription Factor Database. Once hypersensitive regions have been mapped, their DNA sequence is analyzed for potential regulatory elements using the database known as TRANSFAC at <http://transfac.gbf.de/TRANSFAC/index.html>. TRANSFAC is a database of regulatory genomic elements created by the GBF research group.

20 Zinc Finger Design. The well-characterized human transcription factor (Sp-1) is used as a scaffold protein for ZFP design (82, 83). By incorporating amino acid sequence changes into the three recognition helices of this protein, it is possible to design ZFPs to bind to predetermined 9 base pair DNA sequences. 25 Sp1-based zinc fingers can be designed to target most DNA triplets containing 5' KNN, where K is either G or T and N can be any base (61, 63, 84). Hence, target sites for designed three finger proteins are preferably KNN KNN KNN-type sequences, although other sequences can also be targeted (see references cited *supra*). Three finger domains are utilized to recognize 9-10 base pair DNA target sites. In certain circumstances, two three-finger domains can be linked to expand the specificity to 18-20 base pairs, with greatly enhanced affinity (85-87).

30 Once one or more target sites have been selected, appropriate ZFPs are designed. Sequence-specific ZFP are designed and/or selected as described *supra*. In one embodiment, ZFP sequences are obtained from a database containing three-finger ZFP designs that have been characterized for their ability to bind an appropriate DNA target site. Each finger is identified by the complete amino acid

sequence of its recognition helix, and the database contains *de novo* designs based on design rules inferred from mutagenesis experiments, empirically selected fingers from phage display experiments and naturally occurring ZFPs that have been characterized in the literature. Generally, 3-finger ZFPs recognize their 9-10 base-pair target sequences with sub-nanomolar affinities. These affinities are easily an order of magnitude tighter than the affinity of the naturally occurring Sp-1 three-finger domain for its specific target site. In a preferred embodiment for regulation of a gene of interest by a ZFP, the ZFP is targeted to recognize one or more accessible regions close to the transcriptional startsite of the gene.

10        Assembled ZFPs are cloned into pMAL-c2 vectors (New England Biolabs, Beverly, MA), creating maltose-binding protein/ZFP fusions. This permits rapid purification of recombinant ZFPs for characterization. The dissociation constants ( $K_d$ s) of recombinant ZFPs are determined by quantitative electrophoretic mobility shift assay. Briefly, the recognition sequence is incorporated into an end-labeled oligonucleotide and the binding affinities are determined by titrating protein against a fixed amount of oligonucleotide (61, 62). The oligonucleotides have the general format 5'-CATGTATAT-XXXXXXXXX-ATAGAAATGC-3'. In some instances, two 3-finger ZFPs can be linked together to yield a ZFP that recognizes 18-20 bp, to yield higher specificity. Finally, if target sequences are 15 identified, for which no ZFP of appropriate specificity exists in a database, individual fingers are redesigned to recognize those sequences.

20        To evaluate the function of a designed regulatory molecule, a cell-based reporter assay is employed. Accordingly, a ZFP-containing regulatory molecule is inserted into a modified pCDNA3 vector containing, in the following order, a CMV promoter, translation initiation signals, a nuclear localization signal, a multiple cloning site for insertion of a ZFP-encoding sequence, a KRAB domain (64 amino acids), a FLAG epitope, and a polyadenylation signal. A pGL3 reporter plasmid (Promega, Madison, WI) containing four tandem copies of the target sequence between a SV40 enhancer and a promoter driving expression of a 25 luciferase gene, is also inserted into the cells by co-transfection. Luciferase expression is measured in the presence and absence of the vector encoding the regulatory molecule. Appropriate controls are also conducted such as, for example, vector lacking sequences encoding a functional domain and vector lacking sequences encoding a DNA-binding domain.

**Example 6: Tissue culture and animal model systems**

Characterization of the promoter regions of target nuclear receptor genes, as described in Example 4, allows studies to be conducted in tissue culture and 5 animal models. With respect to tissue culture models, regulatory molecules are introduced into human or mouse cell lines using either using lipid-mediated transfection or retroviral transduction. Model cell lines will consist of one that expresses the target gene and one that does not. ZFP expression is verified by Western analysis using, for example, antibody against the FLAG epitope.

10 Chromatin immunoprecipitation (ChIP) methods are used to ascertain specific binding of a designed regulatory molecule to the target promoter. In some cell types, a regulatory molecule may not be functional due to an inactivation of that particular pathway within the cell type. In these situations, various activation domains (see *supra*) are linked to the ZFPs to identify functional activation and 15 repression pathways in each cell type. To understand the function of a particular NR, gene expression profiling is performed in cell lines containing stably integrated sequences encoding a regulatory molecule. For example, disruption of RXR- $\alpha$  and/or inhibition of its expression should compromise all the other NRs that utilize RXR- $\alpha$  as a dimerization partner. Thus, the ability of PPAR- $\gamma$  to 20 upregulate its downstream genes (e.g., PEPCK) should be diminished as well as CAR- $\beta$ 's ability to upregulate its downstream genes (e.g. cytochrome p450). In certain situations, regulatory molecule-encoding sequences are placed in an inducible system to allow for control of their expression. Commercially available cDNA arrays are utilized to obtain the gene expression profiles of stable 25 transfectants. Identification of a class of genes regulated by each nuclear receptor provides information regarding the role of that particular receptor in cellular metabolism.

Transgenic mouse models are used to evaluate NR function by examination of phenotype at a whole animal level and assessment of the 30 physiological role of the nuclear receptors in a whole animal system. For example, an embryonic lethal phenotype is expected in transgenic mice in which an ER- $\alpha$  or HNF4- $\alpha$  repressor ZFP was introduced. Under these situations, the ZFP is introduced into the transgenic mouse under the control of an inducible

promoter, to facilitate viability and allow examination of the role of the nuclear receptor at different stages of development.

### References

- 5 1. Moras, D. and Gronemeyer, H. *Curr Opin Cell Biol* 10 384-391 (1998)
2. Whitfield, G.K., Jurutka, P.W., Haussler, C.A. and Haussler, M.R. *J Cell Biochem Suppl* 110-122 (1999)
3. Barak, Y., Nelson, M.C., Ong, E.S., Jones, Y.Z., Ruiz-Lozano, P., Chien, K.R., Koder, A. and Evans, R.M. *Mol Cell* 4 585-595 (1999)
- 10 4. Wei, P., Zhang, J., Egan-Hafley, M., Liang, S. and Moore, D.D. *Nature* 407 920-923 (2000)
5. Dyson, E., Sucov, H.M., Kubalak, S.W., Schmid-Schonbein, G.W., DeLano, F.A., Evans, R.M., Ross, J., Jr. and Chien, K.R. *Proc Natl Acad Sci U S A* 92 7386-7390 (1995)
- 15 6. Sucov, H.M., Dyson, E., Gumeringer, C.L., Price, J., Chien, K.R. and Evans, R.M. *Genes Dev* 8 1007-1018 (1994)
7. Li, J., Ning, G. and Duncan, S.A. *Genes Dev* 14 464-474 (2000)
8. Korach, K.S. *Science* 266 1524-1527 (1994)
9. Karas, R.H., Hodgin, J.B., Kwon, M., Krege, J.H., Aronovitz, M., Mackey, W., Gustafsson, J.A., Korach, K.S., Smithies, O. and Mendelsohn, M.E. *Proc Natl Acad Sci U S A* 96 15133-15136 (1999)
- 20 10. Krege, J.H., Hodgin, J.B., Couse, J.F., Enmark, E., Warner, M., Mahler, J.F., Sar, M., Korach, K.S., Gustafsson, J.A. and Smithies, O. *Proc Natl Acad Sci U S A* 95 15677-15682 (1998)
11. Byrne, M.M., Sturis, J., Fajans, S.S., Ortiz, F.J., Stoltz, A., Stoffel, M., Smith, M.J., Bell, G.I., Halter, J.B. and Polonsky, K.S. *Diabetes* 44 699-704 (1995)
- 25 12. Shih, D.Q., Dansky, H.M., Fleisher, M., Assmann, G., Fajans, S.S. and Stoffel, M. *Diabetes* 49 832-837 (2000)
13. Rosen, E.D., Walkey, C.J., Puigserver, P. and Spiegelman, B.M. *Genes Dev* 14 1293-1307 (2000)
- 30 14. Gampe, R.T., Jr., Montana, V.G., Lambert, M.H., Miller, A.B., Bledsoe, R.K., Milburn, M.V., Kliewer, S.A., Willson, T.M. and Xu, H.E. *Mol Cell* 5 545-555 (2000)
15. Hayashi, S. and Yoshida, T. *Nippon Rinsho* 58 495-503 (2000)
- 35 16. Lapidus, R.G., Nass, S.J. and Davidson, N.E. *J Mammary Gland Biol Neoplasia* 3 85-94 (1998)
17. Tonetti, D.A. and Jordan, V.C. *J Mammary Gland Biol Neoplasia* 4 401-413 (1999)
18. Pascussi, J.M., Drocourt, L., Fabre, J.M., Maurel, P. and Vilarem, M.J. *Mol Pharmacol* 58 361-372 (2000)
- 40 19. Wiebel, F.F., Steffensen, K.R., Treuter, E., Feltkamp, D. and Gustafsson, J.A. *Mol Endocrinol* 13 1105-1118 (1999)
20. Yamaguchi, S., Murata, Y., Nagaya, T., Hayashi, Y., Ohmori, S., Nimura, Y. and Seo, H. *J Mol Endocrinol* 22 81-90 (1999)
21. Wiebel, F.F. and Gustafsson, J.A. *Mol Cell Biol* 17 3977-3986 (1997)
- 45 22. Staal, A., van Wijnen, A.J., Birkenhager, J.C., Pols, H.A., Prahl, J., DeLuca, H., Gaub, M.P., Lian, J.B., Stein, G.S., van Leeuwen, J.P. and Stein, J.L. *Mol Endocrinol* 10 1444-1456 (1996)
23. Forman, B.M., Tzameli, I., Choi, H.S., Chen, J., Simha, D., Seol, W., Evans, R.M. and Moore, D.D. *Nature* 395 612-615 (1998)
- 50 24. Adachi, M., Takayanagi, R., Tomura, A., Imasaki, K., Kato, S., Goto, K., Yanase, T., Ikuyama, S. and Nawata, H. *N Engl J Med* 343 856-862 (2000)

25. Newmark, J.R., Hardy, D.O., Tonb, D.C., Carter, B.S., Epstein, J.I., Isaacs, W.B., Brown, T.R. and Barrack, E.R. *Proc Natl Acad Sci U S A* 89 6319-6323 (1992)

26. Hu, X. and Lazar, M.A. *Trends in Endocrinology and Metabolism* 11 6-10 (2000)

27. Love, J.D., Gooch, J.T., Nagy, L., Chatterjee, V.K. and Schwabe, J.W. *Biochem Soc Trans* 28 390-396 (2000)

5 28. Fischle, W., Emiliani, S., Hendzel, M.J., Nagase, T., Nomura, N., Voelter, W. and Verdin, E. *J Biol Chem* 274 11713-11720 (1999)

29. Grozinger, C.M., Hassig, C.A. and Schreiber, S.L. *Proc Natl Acad Sci U S A* 96 4868-4873 (1999)

10 30. Huang, E.Y., Zhang, J., Miska, E.A., Guenther, M.G., Kouzarides, T. and Lazar, M.A. *Genes Dev* 14 45-54 (2000)

31. Robyr, D., Wolff, A.P. and Wahli, W. *Mol Endocrinol* 14 329-347 (2000)

32. Kalkhoven, E., Valentine, J.E., Heery, D.M. and Parker, M.G. *Embo J* 17 232-243 (1998)

15 33. Onate, S.A., Boonyaratnakornkit, V., Spencer, T.E., Tsai, S.Y., Tsai, M.J., Edwards, D.P. and O'Malley, B.W. *J Biol Chem* 273 12101-12108 (1998)

34. Onate, S.A., Tsai, S.Y., Tsai, M.J. and O'Malley, B.W. *Science* 270 1354-1357 (1995)

35. Spencer, T.E., Jenster, G., Burcin, M.M., Allis, C.D., Zhou, J., Mizzen, C.A., 20 McKenna, N.J., Onate, S.A., Tsai, S.Y., Tsai, M.J. and O'Malley, B.W. *Nature* 389 194-198 (1997)

36. Yao, T.P., Ku, G., Zhou, N., Scully, R. and Livingston, D.M. *Proc Natl Acad Sci U S A* 93 10626-10631 (1996)

25 37. Choi, H.S., Chung, M., Tzameli, I., Simha, D., Lee, Y.K., Seol, W. and Moore, D.D. *J Biol Chem* 272 23565-23571 (1997)

38. Picard, D. *Nature* 395 543-544 (1998)

39. Baes, M., Gulick, T., Choi, H.S., Martinoli, M.G., Simha, D. and Moore, D.D. *Mol Cell Biol* 14 1544-1551 (1994)

40. Cagnoli, M., Barbieri, F., Bruzzo, C. and Alama, A. *Gynecol Oncol* 70 372-377 (1998)

30 41. Denhardt, D.T. *Ann N Y Acad Sci* 660 70-76 (1992)

42. Nolte, R.T., Conlin, R.M., Harrison, S.C. and Brown, R.S. *Proc Natl Acad Sci U S A* 95 2938-2943 (1998)

43. Bray, P., Lichter, P., Thiesen, H.J., Ward, D.C. and Dawid, I.B. *Proc Natl Acad Sci U S A* 88 9563-9567 (1991)

35 44. Pellegrino, G.R. and Berg, J.M. *Proc Natl Acad Sci U S A* 88 671-675 (1991)

45. Beerli, R.R., Segal, D.J., Dreier, B. and Barbas, C.F., 3rd *Proc Natl Acad Sci U S A* 95 14628-14633 (1998)

46. Beerli, R.R., Dreier, B. and Barbas, C.F., 3rd *Proc Natl Acad Sci U S A* 97 1495-1500 (2000)

40 47. Beerli, R.R., Schopfer, U., Dreier, B. and Barbas, C.F., 3rd *J Biol Chem* 275 32617-32627 (2000)

48. Kang, J.S. and Kim, J.S. *J Biol Chem* 275 8742-8748 (2000)

49. Zhang, L., Spratt, S.K., Liu, Q., Johnstone, B., Qi, H., Raschke, E.E., Jamieson, A.C., 45 Rebar, E.J., Wolff, A.P. and Case, C.C. *J Biol Chem* 275 33850-33860 (2000)

50. Zhang, J. and Lazar, M.A. *Annu Rev Physiol* 62 439-466 (2000)

51. Berg, J.M. *Proc Natl Acad Sci U S A* 85 99-102 (1988)

52. Omichinski, J.G., Clore, G.M., Appella, E., Sakaguchi, K. and Gronenborn, A.M. *Biochemistry* 29 9324-9334 (1990)

50 53. Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T. and Rhodes, D. *Nature* 366 483-487 (1993)

54. Houbaviy, H.B., Usheva, A., Shenk, T. and Burley, S.K. *Proc Natl Acad Sci U S A* 93 13577-13582 (1996)

55. Pavletich, N.P. and Pabo, C.O. *Science* 252 809-817 (1991)

55 56. Pavletich, N.P. and Pabo, C.O. *Science* 261 1701-1707 (1993)

57. Choo, Y., Sanchez-Garcia, I. and Klug, A. *Nature* 372 642-645 (1994)  
58. Choo, Y. and Klug, A. *Proc Natl Acad Sci U S A* 91 11163-11167 (1994)  
59. Choo, Y. and Klug, A. *Curr Opin Struct Biol* 7 117-125 (1997)  
60. Greisman, H.A. and Pabo, C.O. *Science* 275 657-661 (1997)  
5 61. Jamieson, A.C., Kim, S.H. and Wells, J.A. *Biochemistry* 33 5689-5695 (1994)  
62. Jamieson, A.C., Wang, H. and Kim, S.H. *Proc Natl Acad Sci U S A* 93 12834-12839  
(1996)  
63. Rebar, E.J. and Pabo, C.O. *Science* 263 671-673 (1994)  
64. Rebar, E.J., Greisman, H.A. and Pabo, C.O. *Methods Enzymol* 267 129-149 (1996)  
10 65. Wolfe, S.A., Nekludova, L. and Pabo, C.O. *Annu Rev Biophys Biomol Struct* 29 183-  
212 (2000)  
66. Wu, H., Yang, W.P. and Barbas, C.F., 3rd *Proc Natl Acad Sci U S A* 92 344-348  
(1995)  
67. Berg, J.M. and Shi, Y. *Science* 271 1081-1085 (1996)  
15 68. Wolffe, A.P. and Kurumizaka, H. *Prog Nucleic Acid Res Mol Biol* 61 379-422 (1998)  
69. Wolffe, A.P. and Hayes, J.J. *Nucleic Acids Res* 27 711-720 (1999)  
70. Wolffe, A.P. and Guschin, D. *J Struct Biol* 129 102-122 (2000)  
71. Collingwood, T.N., Butler, A., Tone, Y., Clifton-Bligh, R.J., Parker, M.G. and  
Chatterjee, V.K. *J Biol Chem* 272 13060-13065 (1997)  
20 72. Sadovsky, Y., Webb, P., Lopez, G., Baxter, J.D., Fitzpatrick, P.M., Gizang-Ginsberg,  
E., Cavailles, V., Parker, M.G. and Kushner, P.J. *Mol Cell Biol* 15 1554-1563 (1995)  
73. Sadowski, I., Ma, J., Triezenberg, S. and Ptashne, M. *Nature* 335 563-564 (1988)  
74. Thiesen, H.J. and Meyer, W. *Ann N Y Acad Sci* 684 243-245 (1993)  
75. Witzgall, R., O'Leary, E., Leaf, A., Oualdi, D. and Bonventre, J.V. *Proc Natl Acad  
Sci U S A* 91 4514-4518 (1994)  
25 76. Chen, Z., Guidez, F., Rousselot, P., Agadir, A., Chen, S.J., Wang, Z.Y., Degos, L.,  
Zelent, A., Waxman, S. and Chomienne, C. *Proc Natl Acad Sci U S A* 91 1178-1182  
(1994)  
77. Guidez, F., Huang, W., Tong, J.H., Dubois, C., Balirrand, N., Waxman, S., Michaux,  
30 J.L., Martiat, P., Degos, L., Chen, Z. and et al. *Leukemia* 8 312-317 (1994)  
78. Jansen, J.H., Mahfoudi, A., Rambaud, S., Lavau, C., Wahli, W. and Dejean, A. *Proc  
Natl Acad Sci U S A* 92 7401-7405 (1995)  
79. Barroso, I., Gurnell, M., Crowley, V.E., Agostini, M., Schwabe, J.W., Soos, M.A.,  
35 Maslen, G.L., Williams, T.D., Lewis, H., Schafer, A.J., Chatterjee, V.K. and  
O'Rahilly, S. *Nature* 402 880-883 (1999)  
80. Gurnell, M., Wentworth, J.M., Agostini, M., Adams, M., Collingwood, T.N.,  
Provenzano, C., Browne, P.O., Rajanayagam, O., Burris, T.P., Schwabe, J.W., Lazar,  
M.A. and Chatterjee, V.K. *J Biol Chem* 275 5754-5759 (2000)  
81. Tontonoz, P., Hu, E. and Spiegelman, B.M. *Curr Opin Genet Dev* 5 571-576 (1995)  
40 82. Kadonaga, J.T. and Tjian, R. *Proc Natl Acad Sci U S A* 83 5889-5893 (1986)  
83. Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. *Cell* 51 1079-1090 (1987)  
84. Choo, Y. and Klug, A. *Proc Natl Acad Sci U S A* 91 11168-11172 (1994)  
85. Kim, J.S. and Pabo, C.O. *J Biol Chem* 272 29795-29800 (1997)  
86. Kim, J.S. and Pabo, C.O. *Proc Natl Acad Sci U S A* 95 2812-2817 (1998)  
45 87. Liu, Q., Segal, D.J., Ghiara, J.B. and Barbas, C.F., 3rd *Proc Natl Acad Sci U S A* 94  
5525-5530 (1997)  
88. Quelle, D.E., Zindy, F., Ashmun, R.A. and Sherr, C.J. *Cell* 83 993-1000 (1995)

CLAIMS

What is claimed is:

1. A method for regulating the expression of a gene residing in the chromatin of a cell, the method comprising:
  - (a) identifying one or more accessible regions in cellular chromatin associated with the gene;
  - (b) designing a regulatory molecule, wherein the regulatory molecule comprises:
    - (1) a DNA-binding domain targeted to a sequence within the accessible region; and
    - (2) a functional domain; and
  - (c) contacting the regulatory molecule with the cell.
- 15 2. The method according to claim 1, wherein the gene encodes a nuclear receptor.
3. The method according to claim 2, wherein the nuclear receptor is selected from the group consisting of ER $\alpha$ , ER $\beta$ , AR, HNF4 $\alpha$ , HNF4 $\gamma$ , PPAR $\gamma$ , RXR $\alpha$  and CAR $\alpha$ .
- 20 4. The method according to any of claims 1 to 3, wherein the accessible region is identified by virtue of its hypersensitivity to a nuclease.
- 25 5. The method according to any of claims 1 to 4, wherein the DNA-binding domain is a zinc finger domain.
6. The method according to any of claims 1 to 5, wherein the functional domain is an activation domain.
- 30 7. The method according to claim 6, wherein the activation domain is selected from the group consisting of (a)VP16; (b) p65 and (c) functional fragments of (a) or (b).

8. The method according to any of claims 1 to 5, wherein the functional domain is a repression domain.

5 9. The method according to claim 8, wherein the repression domain is selected from the group consisting of (a) KRAB; (b) T; (c) vErbA; and (d) functional fragments of (a), (b) or (c).

10 10. The method according to any of claims 1 to 9, wherein the regulatory molecule is encoded by an expression construct and the expression construct is contacted with the cell.

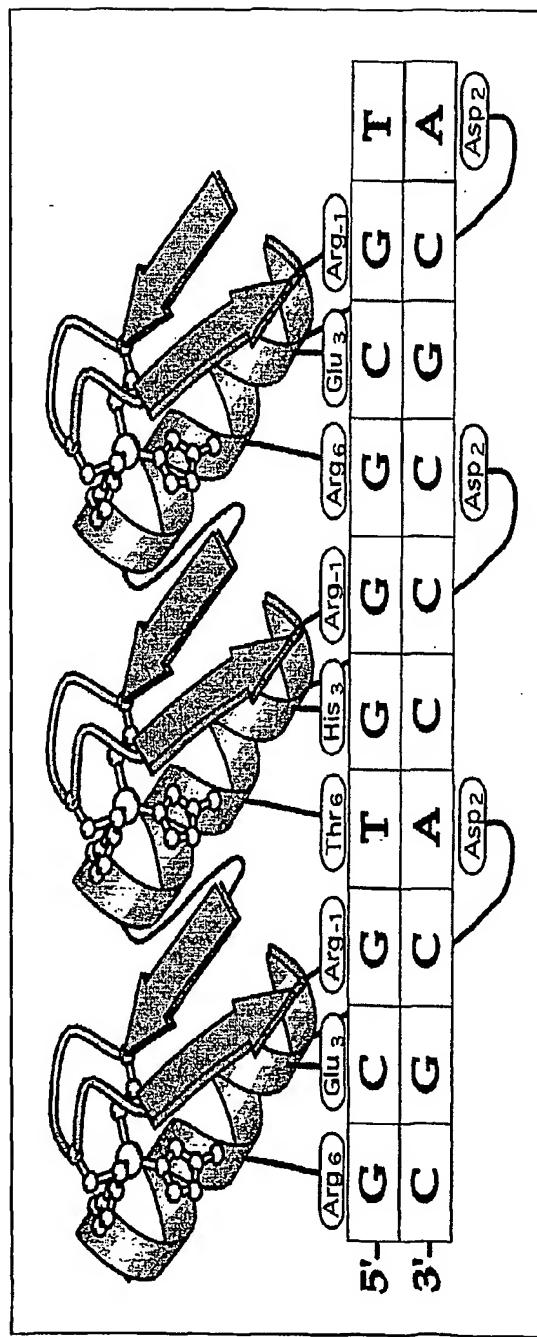


FIGURE 1

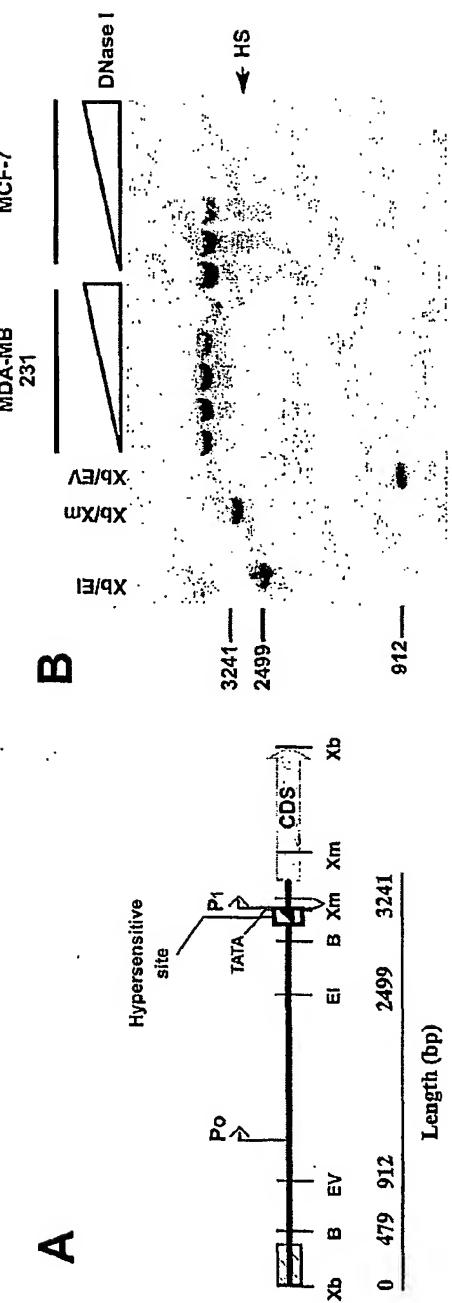


FIGURE 2

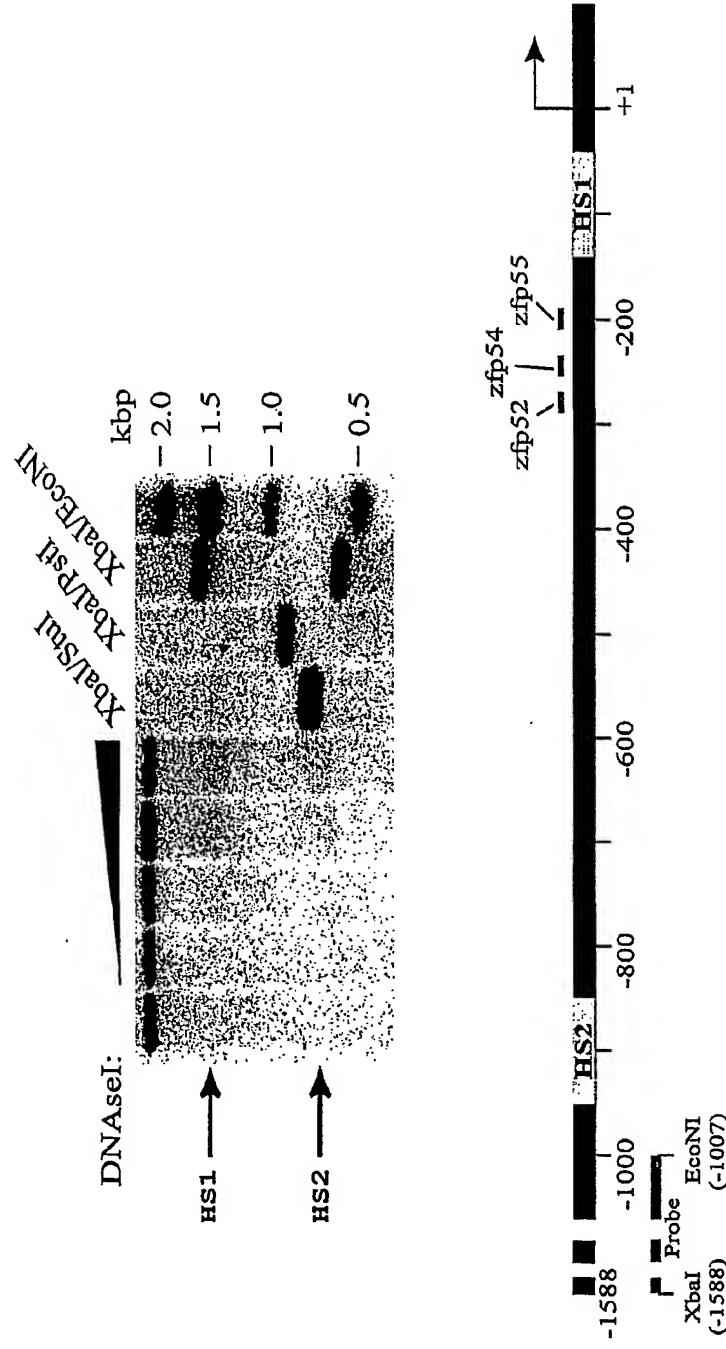


FIGURE 3

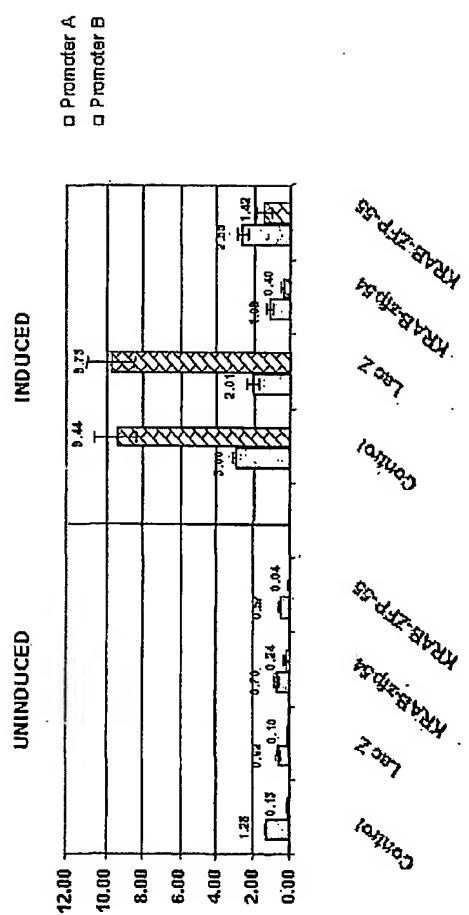


FIGURE 4

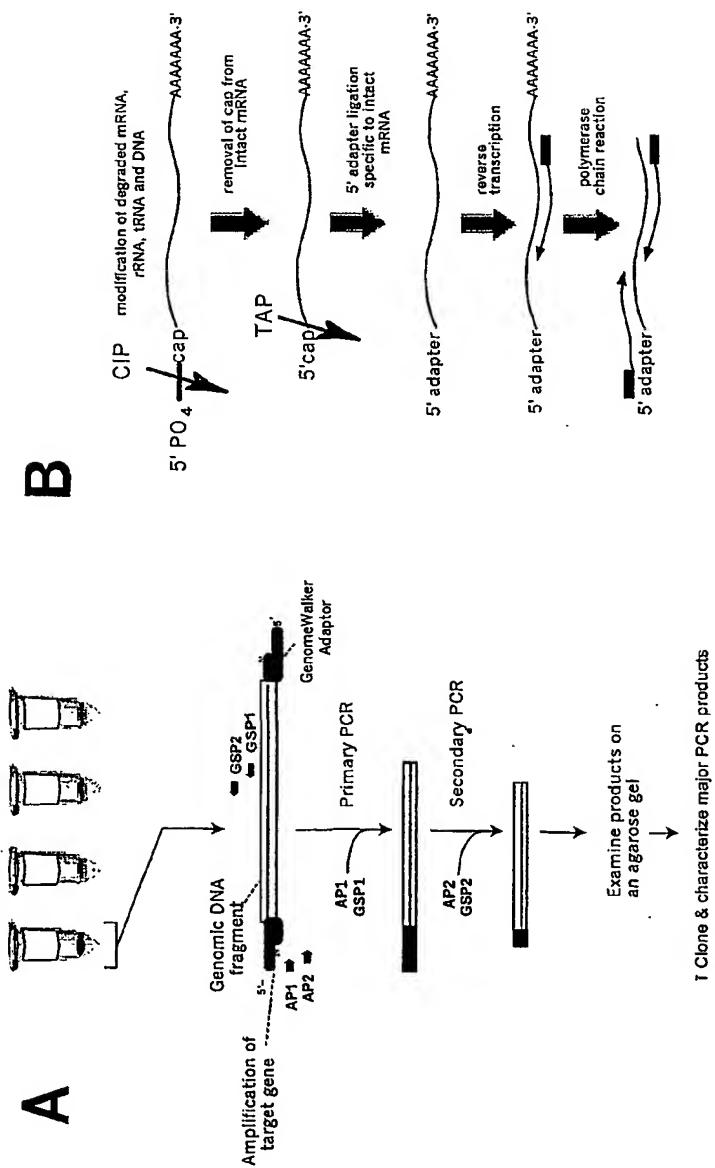


FIGURE 5

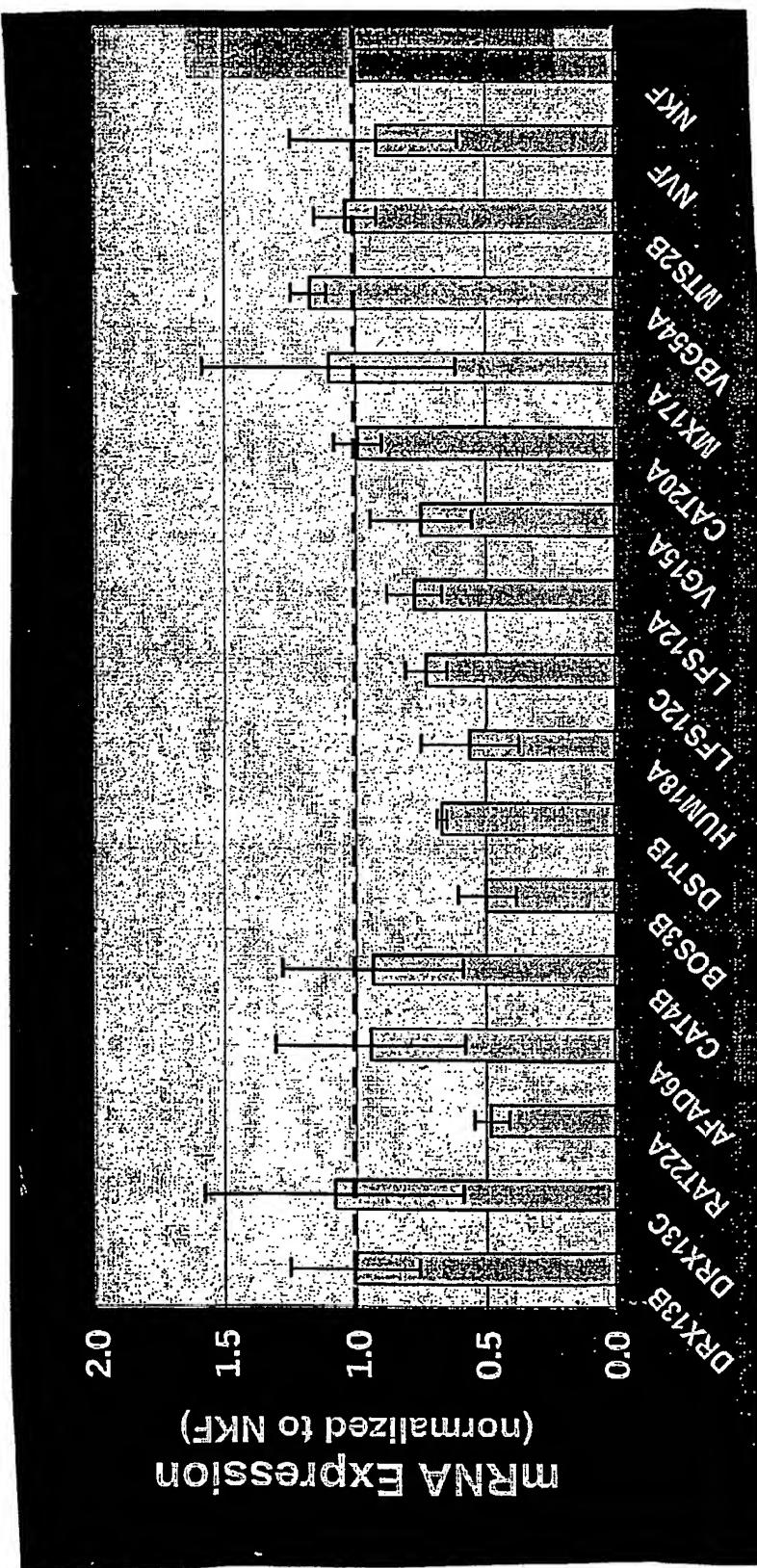


FIGURE 6

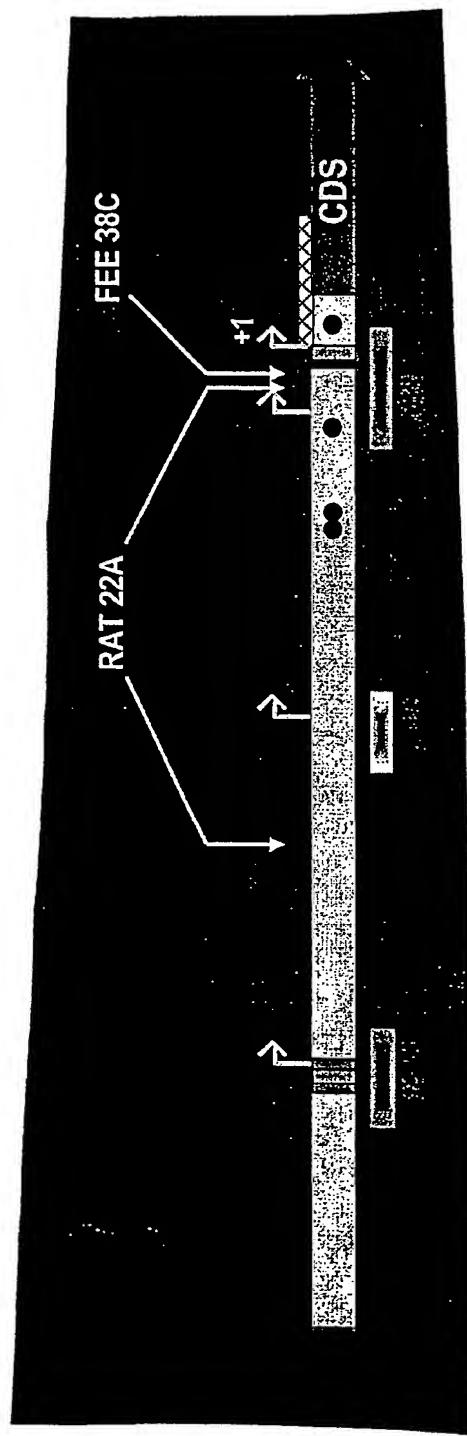


FIGURE 7

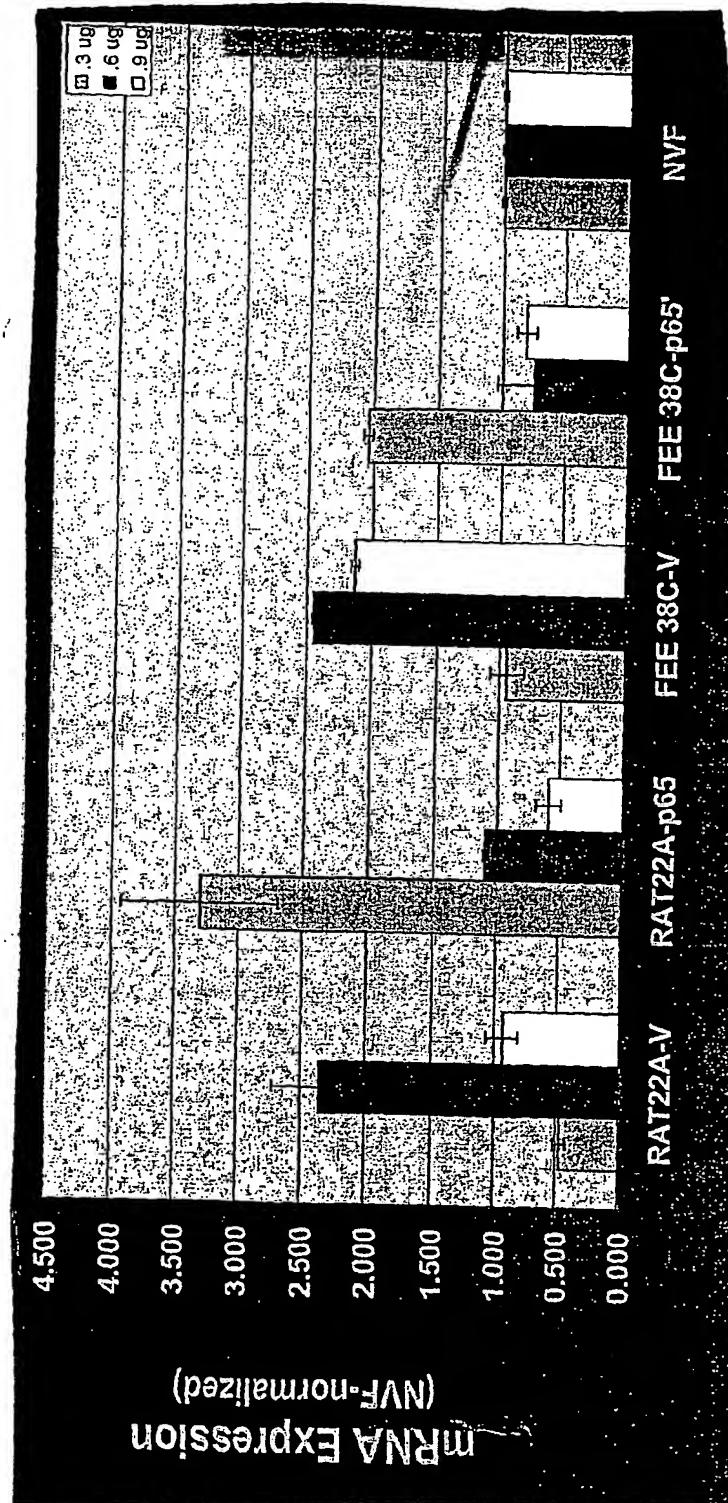


FIGURE 8